



APPLICANT: Dougherty, Brian A  
APPLICANT: Goldberg, Steven L  
APPLICANT: Hofle, Gerhard  
APPLICANT: Mueller, Joachim  
APPLICANT: Reichenbach, Hans  
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or  
FILE REFERENCE: PCT/US 99/23535  
CURRENT FILING DATE: 1999-10-07  
EARLIER FILING DATE: 1998-10-09  
NUMBER OF SEQ ID NOS: 107  
SOFTWARE: Patentm Ver. 2.1  
SEQ ID NO 79  
LENGTH: 1213  
TYPE: PRT  
ORGANISM: *Sorangium cellulosum*  
US-09-413-814-79

Alignment scores:  
Quality: 119.50 Length: 491  
Ratio: 0.660 Gaps: 22  
Percent Similarity: 36.864 Percent Identity: 25.051

Alignment block:

US-09-303-518D-127 x US-09-413-814-79 ..

Align seg 1/1 to: US-09-413-814-79 from: 1 to: 1213

```
18 AGGTCTAACTGCGCCATCGGGGACAGCGGACGAATGATTTATGACG 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
117 ArgArgAlaProAlaAlaArgArgLysHisGlyArgProGlyLeu...Ar 132
68 GGGCCGTCATTCGAGTCCGCTGCTGGTGGCGAAGATATGCGCGTAG 117
   | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
132 GAspleuHisValArgValHisGlyAla..... 141
118 GCGCCCTNGATGAAGTCAAGAAAGCGGATCGGCAAAAAGGCCAAGT 167
   : : : : : : : : : : : : : : : : : : : : : : : :
142 .....AlaGlnGlyArgAspArgProProArg..... 151
168 GCTGTGAAGACAAAGNATCGGGCGGTGTTTACCGCGCGCTTT 217
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
152 .....ArgGlyLeuHisGlyProArgHis 159
218 CAGGCAAAAT...CGCGCCATCATCGCGGCAAAAGCGCTACTTCAG 264
   : : : : : : : : : : : : : : : : : : : : : : : :
159 sGlnProProLeuArgArgArgProGlyGlyProGlyAlaArgAlaLeu 176
265 TCGGT.....CGTATTGCGCGTTGA 284
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
176 AlaLeuLeuArgProValGlyLeuArgValArgAspAlaArgArg 192
285 AGGCAAGCAAGAAATCGAGTTGCAAGCTACCGCGCAAGGCTTGGAA 334
   ||| : : : : : : : : : : : : : : : : : : : : : :
193 ArgArgArgArgArgArgProArgProArgProGlyLeuGlySerGlyAl 209
335 ACTTAAGCGCGGANGAANTNNGCATCTGATCCAAATCCGGTTTGG 384
   ||| ||| ||| : : : : : : : : : : : : : : : : :
209 AlaLeuAlaArgAlaArgGlyAlaArgAlaGlyAspArgValGlyLeu 226
385 ACTGCGCTCGTANCCGCTTCAGCAAAATCCCTGCCGT...CGATGC 431
   ||| : : : : : : : : : : : : : : : : : : : : : :
226 roGlyAlaAspArgLysAspAlaHisGlyArgValProArgArgGlyArgPro 242
432 CGAGCCGTT.....CGCCAT...CTTCGTCATCGCATGACGA 466
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
243 GlyAlaValAlaValAlaProProArgHisAspArgLysArgLeuAspProAl 259
467 CCAATCCGTTNCGCGACACCTGTGTTGATCAAAAGAACCGCGCAN 516
   : : : : : : : : : : : : : : : : : : : : : : : :
```

```
259 acLAlaProArgSerHisProArgGlyLeuProArgAlaProArgArg 276
517 GATTTCAGACGANGTTCGTGATTTAG.....CGTTGAC 554
   : : : : : : : : : : : : : : : : : : : : : :
276 LeProArgArgArgArgArgGlyAlaAspLeuValAspArgProArg 292
555 CGAGCGTAAATTCATGTTGTGTAAAGGACGTGGCGACAGCGCGCTG 604
   ||| : : : : : : : : : : : : : : : : : : : : : :
293 ArgGlyArgArgProGlyVal..... 299
605 AAAAGTCCCAATCGAACAACATGATTCGGCGCGCGCATCGGCC 654
   : : : : : : : : : : : : : : : : : : : : : :
300 .....AlaGlnHisPro.....LeuArgProProAla.....A 309
655 GCTTGAATGCGACGACATTCATTCATTCAGCGCGTCCGTCAAA... 701
   ||| ||| : : : : : : : : : : : : : : : : :
309 rGlyProAlaHisLeuArgAlaArgArgGlyAlaGlyAlaValSerAsp 325
701 ..... 701
326 ArgGlyProArgArgArgProHisArgArgAspArgArgAlaArgTll 342
701 ..... 701
342 eLeuAlaArgArgGlyAlaAspProGlyAlaValProGlyAlaProHisA 359
702 .....CAAAACCGTTTGACCATCATTCATTCAGAGTGT 734
359 sPArgArgAlaAlaValGlnAspArgArgProGlyProLeuLeuArgArg 375
735 AATGCGCATCGAGCGTTTGTTCACACAGCGCTCT..... 770
   : : : : : : : : : : : : : : : : : : : : : :
376 ArgAspHisArgArgAlaAlaArgAlaHisArgProSerGlyAlaAspProArg 392
771 .....GAACACCGACGCGGTGATTCCTTGGGTGTTCTCAAGTCAACAA 816
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
392 gLeuProHisArgAlaArgArgAspArgGly.....ArgProAlaG 406
817 CCAAGCGCTCTGCGTACCGCTTTGGGTGCGAAAGATGCGCAATTCATGCG 866
   ||| : : : : : : : : : : : : : : : : : : : : : :
406 lAlaLeuProLeu..... 409
867 GGGCGAATTCGTTGACGACAGACACCGCGTATTCGCGTTCGATTTGA 916
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
410 GlyArgAlaGlyArg.....GlyGlyGly 418
917 ACGCGCGGATTCACAAAGCGCGGACAGATTTATTCGACGCTACCAAT 966
   ||| : : : : : : : : : : : : : : : : : : : : : :
418 uAspArgSerValGlyArgGlyAla.....ProGlyA 429
967 CAGATTCCTGTTATCGAAGAGCGCGCAAGAGAGCTGTTGGCTGGGT 1016
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
429 rGlyAlaArgArgArgArgArgArgArgArgArgArgAlaProArgLeu... 444
1017 TGGCGCGGACGCGCAAAATCTCATGACGCGGATTCGCGCATTT 1066
   ||| : : : : : : : : : : : : : : : : : : : : : :
445 .....ArgPro... 446
1067 TCCTGAACAACAACTTTCAAGTTTCAGACAGACCGCTCAAGCGTGGGAC 1116
   : : : : : : : : : : : : : : : : : : : : : :
447 .....GlnGlnAlaAlaGlyValHisAspProArgGlyGlyArgArg 461
1117 CGCGCGCANGTCCGATTCGATTCAGAGCGCGCTAAATGCGCGTACGAT 1166
   ||| ||| ||| : : : : : : : : : : : : : : : : :
461 ro...ProGlyAlaAlaAlaGlnArgGlnArgGlnGlyArgProArgArg 476
1167 CCGTACCTACCTGCTTTGCGCGATTTAATGCTGG..... 1202
   ||| ||| ||| : : : : : : : : : : : : : : : : :
477 ProAlaGlySerArgArgGlyArgAlaAlaArgArgGlyArgArgProAl 493
1203 .....CGATACGACAGCGCGGACAGATTCGCTGGTGTTC 1236
   : : : : : : : : : : : : : : : : : : : : : :
493 aAspArgAspArgAlaAlaHisArgLeuGlyAlaArgGlyLeuAlaG 510
```



627 .ArgAlaGlyProAlaPheProAlaAlaGlyAlaProGlyAlaValArg 642  
seq\_name: /cgn2\_6/ptodata/1/lae/6A\_COMB.pep:US-08-997-897-2  
seq\_documentation\_block:  
Sequence 2, Application US/08997897C  
Patent No. 6114514  
GENERAL INFORMATION:  
APPLICANT: SRIVASTAVA, RANJANA  
APPLICANT: KUMAR, DEEPAK  
APPLICANT: SRIVASTAVA, BRAH SHANKER  
TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS SPECIFIC DNA FRAGMENT  
FILE REFERENCE: 001469-7  
CURRENT APPLICATION NUMBER: US/08/997, 897C  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 430  
TYPE: PRT  
ORGANISM: Mycobacterium tuberculosis  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (4)  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (6)  
NAME/KEY: UNSURE  
LOCATION: (20)  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (29)  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (54)  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (64)  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (69)  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (89)  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (99)  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (114)  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (119)  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (129)  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (159)  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (169)  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (182)  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (185)  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (219)  
FEATURE:

NAME/KEY: UNSURE  
LOCATION: (259)  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (269)  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (291)  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (323)  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (339)  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (349)  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (356)  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (366)  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (410)  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (421)  
US-08-997-897-2

alignment\_scores:  
Quality: 112.00 Length: 441  
Ratio: 0.615 Gaps: 24  
Percent Similarity: 41.270 Percent Identity: 23.583

alignment\_block:

US-09-303-518d-127 x US-08-997-897-2 ..

Align seg 1/1 to: US-08-997-897-2 from: 1 to: 430

44 GACCGAGCAAGTCATTATGACGGGCGGTCATTA.....CCGA 84  
|||||::: ::::: |||||::: ||:::  
34 AspArgGluIleAspGlyGluSerGlyProGlyTyrArgProProPhe 50  
85 GTGCGCTTCCTTGGCGAAGATATGCCGGTATGCGCCCTTGATGAAGT 134  
:::|||||::: ||||| |||||  
50 GSerHisCys\*\*\*Ser.ArgProCysArg.....LysArg 61  
135 CAAGGAAGGCGATGCCGTCAAAAAAAGGCC.AAGTGC..... 169  
::: |||||::: |||||::: |||||::: |||||  
62 ArgTyr\*\*\*ArgCysArgAsp\*\*\*ArgSerArgCysTyrCysHisProG 78  
170 .....TGTTCGAAGCAAAAAAAGNATCCGGCGTGT 200  
78 YTyrSerCysArgTyrTyrCysHisProArg\*\*\*SerCysArgGlySer 95  
201 GTTACCGCGCGCGCGTTCAGGCAAAATCGCCGATCATCGCGCGAAA 250  
|| ||||| ||||| ||||| |||||  
95 Ys.TyrArgArg\*\*\*SerArgArgCysSerCysSerCysArgTyr 111  
251 AGCGGCTACTTCAGTCGCTCGTATGCGCGTTGAAGCAAGCAAGATC 300  
: ||||| ||||| ||||| |||||  
111 rProCys\*\*\*Ser.....ArgArgArg\*\*\*G 120  
301 GAGTTCGAACGCTACGCGCCCGAAGCGTTCGCAAACTTAAGCGCGANGA 350  
|| ||||| ||||| ||||| |||||  
120 lYcysProCysCysSerCysGlnHis\*\*\*GlyCysArgTyrCysArgTyr 136  
351 ANTNGNNGCAATCTGATCCAAATCCGGTTGTGAGACTGCGCGCTCGTANCC 400  
::: ||||| ||||| ||||| |||||  
137 ProGlySerArgTyrProSerArg.....Cys.....Pr 147

```

401 GTCGCTTCAGCAAAATCCCTCCCTGCATCCGACGCCGTTGCCATCTTC 450
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
147 oserleuarg.....CysArgArgPheArgCysProArg***ArgC 161
451 GTCATATCGATGACACCAATCCGCTGCGGACGACCCCTGTGTGTGAT 500
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
161 ysglnargtyltpcysProasn**ThrGlyArgCysCysArgCysPro 177
501 CAAGAAGCGCGNCGANGATTCAGACGANGTTCGTGATTTGAGCCGTT 550
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
178 SerSerSerArg***GlnTyr...SerProAlaGlyCysArgArgGln 193
551 TGACCGAGCGTAATAATCATGTGTGTGTAAGGACGTCGCGGAGCGTCCG 600
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
193 rlaaArgCys...ArgCysCysCysCysArgCysTyrPheArgGlyCysC 209
601 TCTGAAAATGCTGCGCAA...CATCGAAACACATGAAATCGCGCGCGCA 647
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
209 ys.....CysCysArgCysTyrPheGlnSerLeuGly***SerArgProArg 223
648 TCCGCGCGGTTTGTAGTGACGACCATTCATTCATGAGCGCGTCCGCTG 697
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
224 SerArg.....SerArgArgCys.....SerArgArgCys 229
698 CAACAACAAACCGTTTGGACCATCATATTCAGATGTAATTCATTCGCA 747
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
229 sSerArgArgArgPheGlnAsnArgCysCysArgSerArgGlyPheArg 246
748 CGTTTGTTCGACACGCGCG...TCTGACACAC.....* 776
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
246 leArgCysCysSerPheProGlyPheArgAsnArgHis***IleLeuArg 262
777 .....CGACGCGCTGATTG 790
263 CysPheHisCysArgTyr**SerCysArgArgCysArgCysArgArgCys 279
791 CTTTGGGAGGTTCTCAAGTCAA...CAACACACGCTTCTGCTACCGCTT 837
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
279 sPheGlyCysArgGlyCysArgCysGlnGlyCys***SerHisArgArgP 296
838 TTGGGTGCGAAAGATTCGCAAT...TACTCGCGGCGCAATTCGTTGCGC 884
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
296 heArgCysArgGlyCysCysAsnCysArgCysTyrArgCys..... 309
885 AGACACCGCGGATTCGCGTTCGATTCGACGCGCGCATTCACACAG 934
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
310 .....ArgGlyCysSerArg 314
935 GCGCGGACGATTAATTGGG.....ACGC 957
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
314 gatgProGlyLeuProGlyArgAsp***ArgProValGlyHisArgLysI 331
958 TACCACATCATGATTCGCTTACGAAGAGCGCGCACCAAGAGAGCTGTT 1007
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
331 leProHisCysCysPheArgCys***ArgSerProArgSerArgProAla 347
1008 CGGCTG.....GGTTCGCGCGGACCGCGACCAATTCCTCATCAGCGC 1048
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
348 leu**TyrProProGlySerCys***ThrAsnProIleArgCysCysPr 364
1049 GTACGACCGCTCGGCAATTCCTGAAAAACAAACTTCAAGTTCACAGCA 1098
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
364 oSer***SerArgProIleProAlaArgProAlaGlyLeuPro.....GlyA 379
1099 GCGGTCAACGCTGGCGACCGCGCGGCGCGCATTCGCTGACTTACAGCGC 1148
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
379 rGserTyrArgTyrProProThrLysSerGly.....Argser 391
1149 CGTATGCGCGGCTAGACATCCTGCTACCGCTTTCGCGGATTTAATCG 1198
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
392 GlnAsnCysTyrPheHisArg.....SerSerGlySerArg 402

```

```

1199 TCGGCGATACGACAGCGC 1217
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
402 gThrArgTyrArgHisArg 408
seq_name: /cgn2_6/plodata/1/laa/5b_COMB.pep:US-09-156-836B-2
seq_documentation_block:
; Sequence 2, Application US/09156836B
; Patent No. 6242585
; GENERAL INFORMATION:
; APPLICANT: Srivastava, Ranjana
; APPLICANT: Srivastava, Brahm Shanker
; TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS SPECIFIC DNA FRAGMENT
; FILE REFERENCE: U 011876-4
; CURRENT APPLICATION NUMBER: US/09/156,836B
; CURRENT FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 08/997,897
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (4)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (6)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (20)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (29)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (54)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (64)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (69)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (89)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (99)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (114)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (119)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (129)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (159)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (169)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (182)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (185)

```

```

1  ? OTHER INFORMATION: amino acid has not been identified
2  ? NAME/KEY: UNSURE
3  ? LOCATION: (219)
4  ? OTHER INFORMATION: amino acid has not been identified
5  ? NAME/KEY: UNSURE
6  ? LOCATION: (259)
7  ? OTHER INFORMATION: amino acid has not been identified
8  ? NAME/KEY: UNSURE
9  ? LOCATION: (269)
10 ? OTHER INFORMATION: amino acid has not been identified
11 ? NAME/KEY: UNSURE
12 ? LOCATION: (291)
13 ? OTHER INFORMATION: amino acid has not been identified
14 ? NAME/KEY: UNSURE
15 ? LOCATION: (323)
16 ? OTHER INFORMATION: amino acid has not been identified
17 ? NAME/KEY: UNSURE
18 ? LOCATION: (339)
19 ? OTHER INFORMATION: amino acid has not been identified
20 ? NAME/KEY: UNSURE
21 ? LOCATION: (349)
22 ? OTHER INFORMATION: amino acid has not been identified
23 ? NAME/KEY: UNSURE
24 ? LOCATION: (356)
25 ? OTHER INFORMATION: amino acid has not been identified
26 ? NAME/KEY: UNSURE
27 ? LOCATION: (366)
28 ? OTHER INFORMATION: amino acid has not been identified
29 ? NAME/KEY: UNSURE
30 ? LOCATION: (410)
31 ? OTHER INFORMATION: amino acid has not been identified
32 ? NAME/KEY: UNSURE
33 ? LOCATION: (421)
34 ? OTHER INFORMATION: amino acid has not been identified
35 US-09-156-830B-2

```

alignment_scores:	
Quality:	110.00
Ratio:	0.625
Percent Similarity:	39.462
alignment_block:	
Length:	446
Gaps:	23
Percent Identity:	23.767

Align seg 1/1 to: US-09-156-836B-2 from: 1 to: 430

```

      ..CGGAGCAGACAGATATTATGAACGGCCCGCTCATTA.....CGCA 84
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
34 AsparitgguIleasprgiYusergIyProglIyTyrAtgProPtoPrAr 50
      85 GTTCGCGTGTTGCCGAAAGATATGCCGGTATGCGCCCTGTGANGAAGT 13
      gserHisCys***Ser.ArgProCysArg.....LysArg 61
50
135 CAAAGAAGCGCATGCCGTCAAAGAACGCC..ANGTCC.....169
62 ArgTyr***ArgYsaIrsAspr**ArgserAtgcYstrPCysHisProGl 78
170 .....TGTTAAACAACAAAAGNATCCGGCGCTGT 200
78 YTYrSerCysAtgYTYrTPCYshIsProArig***SerCysArIgCystrpc 95
201 GTTTACCAGCGCCNGTTTTCAGCGAAANAATCGCCCATTCATCGCGCGAAN 250
    || |||||:::|||||:::|||||:::|||||:::|||||:::
95 ys..TYrArgArg***SerHisArgIrcysSerCysProCysArgse 111
251 AGCGGCTACTTAGTGCgtTcGTATtGCCTtGAAGCAACAGCACGAANC 300
    |||||:::|||||:::|||||:::|||||:::|||||:::
111 rProCys***Ser.....ArgIdArgYat***G 120
301 GAGTTCGAACGCTACGGCCCGCGAAGCGTTTGCAAACTTGAAGCGCGCANCA 350
    ::::::: |||

```

120 lrcyspProcysCysSerCysGlnHis\*\*GlyCysArgTlyCysArgTly 136  
351 ANTNNGNCATATTCATTCATTCGCGTTTGGACCTCGCGTGNCC 400  
137 ProGlySerArgTlyProSerSerArg.....Cys.....Pr 147  
401 GTCGGTTCAGCAAAATCCCTGCCCGTCATGCCAGCCGTTGCCATCTTC 450  
147 oSerLeuArg.....CysArgArgPheArgCysProArg\*\*\*ArgC 161  
451 GTCATTCGCATGGACACCAATCCGCTGNCGGCAACACCCTGGTTGTGAT 500  
161 ysglnArgTlyTrpCysProasn\*\*\*ThrGlyArgCysCysArgCysPro 177  
501 CAAGAAGACCGGNGANGATTTACAGCAGANGTNGCTGGATTCAGCGCTT 550  
178 SerSerSerArg\*\*\*GlnTyr.....SerProAlaGlyCysArgArgTh 193  
551 TGACCGAGCGGTAAATTCATGCTGTGTGAAGGACGCTG..... 587  
193 rAlaArgCys...ArgCysCysCysArgCysTrpArgThrArgCysC 209  
587 ..... 587  
209 yscCysArgCysTrpArgInSerLeuGly\*\*\*SerArgProArgSerArg 225  
588 .....CGCAGACGTCGCGCTGTGAATATGCTGCCAATCG 622  
226 SerArgArgCysSerArgArgArgArgPheGlnAsnAlaArgCysArgSerArg 242  
623 AAACACATGATGATGGCGCGCGCCGACATCCGACGTTTGATGGACACGAC 672  
242 gGlyPheArgTlleArgCysCysSerPheProGlyPhe.....A 255  
673 ATTCAATTCATTCAGCGCGCTGGGTGAACCAAAACCGTTTGACATCA 722  
255 rGlnArgHis\*\*TlleAlaArgCysPheHisCysArgTly...\*\*\*Ser 270  
723 TTATCAGATGATATTCATTCGATGCACGTTTGTTCGCAACAGCCGCTGTA 772  
271 CysArgArgCysArgCysProArgCysPheArgCys..... 282  
773 ACACCGACCGCGATGATTCGTTGGGTGGCTGTCGCAAGTCACAAACACCG 822  
283 ...ArgGlyCysArgCysGlnGlyCys\*\*\*SerHisArg..... 294  
823 CTCCTGGATCCGTTTGGGTGCGCAAGATTCGCAAT...TACTCGCG 869  
295 .....ArgPheArgCysArgGlnGlyCysAsnCysArgCysTrp 307  
870 CGAATTCGTTGACGAGACAAACCGCGGATTTCCGCGTTCGGATTCAGAC 919  
308 ArgCys..... 309  
920 CGCGCATTCACAAAGCGCGCAGCATTAATTATTTGG..... 953  
310 .ArgGlnCysSerArgArgProGlyLeuProGlyArgAsp\*\*\*ArgPro 326  
954 .....ACGTACCAACATACAGATTTCCGTTATCGAAGAGCGCG 992  
326 alGlyHisArgTlyLleProThrCysCysPheArgCys\*\*\*ArgSerPro 342  
993 CAGCAAAAGACTGTTCCGCTG.....GGTTGCGCGGACGCGGAC 1033  
343 ArgSerArgProAlaLeu\*\*\*TrpProProGlySerCys\*\*\*ThrAsnPr 359  
1034 AATTAATTCATCAGCCGATACACCCCTGCGCATTTCTTGAACAAACAACT 1083  
359 oLleArgCysCysProSer\*\*\*SerArgProLleProAlaArgProArgL 376  
1084 TTCAAGTTCAAGACAGACGCGTCGACAGGTTGGACCGCGGCAATGGCGCAT 1133  
376 euPro.....GlyArgSerTyrArgTlyProProThrLlyysSerGly 389

```

1134 TGGTACTACGAGCGCGTAATCCGCTAGACATCGTCCCTACCTGCTTT 1183
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
390 .....ArgSerGlnAsnCysTrpHisArg..... 397

1184 TCGCGGATTTAATGTCGCGCATACCGACAGCCG 1217
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
398 .SerSerGlySerArgThrArgTyrArgHisArg 408

seq_name: /cgn2_6/prodata/1/aa/backfiles1.pcp:5223423-4

seq_documentation_block:
; Patent No. 5223423
; APPLICANT: FRANCHINI, GENOVEFFA, WONG-STAL, FLOSSIE;
; GALLO, ROBERT
; TITLE OF INVENTION: CHARACTERIZATION OF REPLICATION COMPETENT
; HUMAN IMMUNODEFICIENCY TYPE 2 PROVIRAL CLONE HIV-2 SBL/ISY
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION NUMBER: US/07/331,212
; FILING DATE: 03-31-1989
; SEQ ID NO: 4
; LENGTH: 3080
5223423-4

alignment_scores:
      quality: 104.50      length: 423
      ratio: 0.562      gaps: 25
Percent Similarity: 42.956      Percent Identity: 21.940

alignment_block:
US-09-303-518D-127 x 5223423-4

Align seg 1/1 to: 5223423-4 from: 1 to: 3080

195 CGTGGTGTTCACCGCCGCTTCAGCAAAATCCGCCATCCATCCGCG 244
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
256 ArgHisLeuValLeuAlaArgArgArgGluSerGluArgTyrArgSer 272
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
245 GCGAAACCGCGT...ACTGATGTCGT...CGTATGCGCTTGA 285
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
272 slyssnsrArgGluThrSerSerGlyArgAsnArgAsnCysArgGlu 289
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
286 GCGACAGCAAGATCGAGTTCGACGCTACGCGCCGACGCTTGCGCAA 335
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
289 snAlaIstYtLysThrAsnSerThrThrArgGluArg.....GlyLys 303
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
336 CTTAACCGCGCGGANGAANTNNNGCAATCTGATCCATCCGCTTGTGA 385
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
304 LeuProAlaGlaThrAsnArgArgGln..... 312
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
386 CTGCGCTGCGTANCCGCTCCGTCAGCAAAATCCCTGCCGTCGATCCGAG 435
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
313 .....LeuCysProSerAlaAlaGluSerProAsnProLysLeuG 327
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
436 CC...GTTGCCCATCTTCCTCAATGCGATGACACCAATCCGCTGCG... 479
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
327 LysValSerArgGlyGluLysValArgGlyArgSerSerAlaGlyTle 343
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
480 .....GG 481
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
344 SerGlyThrLeuArgArgLeuHisAlaLeuTyrSerAsnAlaLeuCys 360
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
482 CAGACCTGTGCTTGATCAAGAGCGCGANGATTCAGACGAGCT 531
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
360 yArgProSerSerSerAspAlaAsnAsnGlnArgAsnTyrArgSerSer 377
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
532 NTGCGGATTTGAGCCGCTTGACCGAGCGTAAATCATGTGTGAAGC 581
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
377 rgleugly..... 379
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
582 AGCTGGCGCAGAGCTGCCGCTCTGAAAATGCTGCAACATCGAAACATG 631

```

```

380 .....CysThrThrSerAsnThr... 385
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
632 AATTCGCGCGCCCGCATCCGCGGCTTGTAGT...GCGACGACATTT 675
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
386 .....ArgProLeuThr.SerGlyAlaAlaGlnArgSerThrArgile 399
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
676 CATTTATTTAGCCCGCTGCGTGCACAAACAAACCGTTTGACATCAATT 725
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
400 HisSerArgAspAsnLysHisSerArgGlyThr..... 410
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
726 TCAAGATGTAATGCGCATCGAGCGTTTGTTCACACAGCCGCTTGACA 775
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
411 .....AspArgMetAspValAlaArgLysSerCysThrSerArgLysHis 426
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
776 CCGAG.....CGCGTGTTCCTTGG 795
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
426 euGluMetAspProAspArgThrAlaGluValCysGlnAsnValGlnSer 442
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
796 GGTGGTCTCAGTCAACACACCGCTTTCGCTACCGTT..... 837
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
443 AsnGlnHisSerArgHisLysThrArgThrLysArgValAlaProLysLe 459
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
838 .....TTGGGTGCGAAAGATATGCAAAATTAATCGCGG 868
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
459 uCysGlyLLeuGlnLysLeuLysGlyArgThrAspArgCysSerSerG 476
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
869 GCGAATTCGTGACGACGAC.....AACCGCTGATT 900
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
476 LglLLeuAspAspProAspAlaAlaSerAlaIleGluProArgLeuVal 492
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
901 TCCGGTTCGATTGACCGCGCGATTACACAAAGCGCGGACGAT..... 945
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
493 Ser.....ThrLysGlyThrArgAspLysLe 501
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
946 .TATTTGGAGCGCTACCAATTCGATTCCGTTATCGAAGAGCGCGCA 994
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
501 rTyrLeuArgArgAspAlaAsnArgLeuSer.....ArgA 513
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
995 GCAAGAGCGTTCGCTGCGCTGCGCGCGACGCG.....GAGCA 1034
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
513 spArg.....TrpThrArgProGluGlnThrAsnGlyArg 525
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1035 ATACTCCATCAGCGCGTACGACGCTGCGCATTTCT..... 1070
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
526 SerLeuLysArgGlyHisAlaThrSerProTyrProLecysSerSer 542
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1071 .....GAAAAA...CAAACTCTCAAGTTACGACAGCGCT..... 1103
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
542 oThnGluLysGlyAsnValLeuGlnLeuThrLysGlyArgAlaLeuGly 559
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1104 .....CAACGCTGG.....CGA 1115
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
559 ysrHrMetProSerAlaLysThrGlyLeuLeuGlnMetTrpLeuValArg 575
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1116 CCGCGCCATGCT.....GCCGATTGCTACTACGAGCG..... 1148
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
576 ThrHisHisGlyLysLeuProArgThrGlyThrPhePheArgAlaTrpTh 592
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1149 .....CGTAATGCCCGCTAGACATCTGCTTACCGCTGTTGGCGCAT 1191
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
592 rMetGlyLysGlyAlaProGlnLeuProArgGlyProLysPheAlaGlyA 609
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1192 TTAATCGTCGCGCATACCGACGCGCGCAAGCATGGGTTGCTTGA 1238
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
609 lAsnThrAsnSerThrProAsnLysSerSerArgLysProThrGly 624
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq\_name: /cgn2\_6/prodata/1/aa/5A\_COMB.pcp:US-08-046-585-5

seq\_documentation\_block:  
; Sequence 5, Application US/08046585  
; Patent No. 5453362  
; GENERAL INFORMATION:



```

1  APPLICANT: Lamarco, Kelly
2  APPLICANT: Wilson, Angus
3  APPLICANT: Herr, Winship
4  TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN
5  TITLE OF INVENTION: HOST CELL FACTOR
6  NUMBER OF SEQUENCES: 15
7  CORRESPONDENCE ADDRESS:
8  ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
9  STREET: 4 Embarcadero Center, Suite 3400
10 CITY: San Francisco
11 STATE: CA
12 COUNTRY: USA
13 ZIP: 94111-4187
14
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: Floppy disk
17 COMPUTER: IBM PC compatible
18 OPERATING SYSTEM: PC-DOS/MS-DOS
19 SOFTWARE: Patentin Release #1.0, Version #1.25
20 CURRENT APPLICATION NUMBER:
21 APPLICATION NUMBER: US/08/046,585
22 FILING DATE: 12-Apr-1993
23 CLASSIFICATION: 433
24
25 ATTORNEY/AGENT INFORMATION:
26 NAME: Osman, Richard A
27 REGISTRATION NUMBER: 36,627
28 REFERENCE/DOCKET NUMBER: A-57503-1/RAO
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: (415) 781-1989
31 TELEFAX: (415) 398-3249
32
33 INFORMATION FOR SEQ ID NO: 5:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 2035 amino acids
36 TYPE: amino acid
37 STRANDEDNESS: single
38 TOPOLOGY: linear
39
40 MOLECULE TYPE: peptide
41
42 US-08-046-585-5

```

alignment_scores:	
Quality:	100.50
Ratio:	0.333
Percent Similarity:	40.379
alignment_block:	
Length:	634
Gaps:	32
Percent Identity:	20.662

US-09-303-518D-127/rev x US-08-046-585-5

CO: 05-08-046-585-5 FROM: 1 TO: 2035

```

1293 CAGCCTTCCTTCACGACGTTTCACACCTTACGCAACAGCGGGCCMTA 1294
      :::::||||:
747 LysrThrIleuLeuGlyIleSerVal.....SerProSe 759
      :::::||||:
1293 TTGGATTGTCGGGGCAGACAGCTGCACAAACCGAGCTCTTCTGT 1244
      ||||| ||| :::: |||::: |||:
759 rThrThrLysProGlyThrThrThrIleLysThrIleProMetSera 776
      :::::||||:
1243 CCATATTCACAGACACCCAAATGCTTCGGCGCTGTGCGTACCGCAGATT 1194
      ||| ||| ::::: |||
776 LalleIleThrGlnaIaGlyAlaThrGlyValThrSerSerProGlyIle 792
      |||||
1193 AAATCG..... 1188
      |||||
793 LysSerProIleThrIleIleThrThrLysValMetThrSerGlyThrG1 809
1187 .....CCCAAAAGCAGGCGTAGCGCAGA 1166
      ||| :::::||||:
809 yAlaProAlaLysIleIleThrAlaValaProLysIleAlaThrGlyHis 826
1165 TGCTCAGCGCAGATTACGGCG..... 1146
      :::::||||:||||:
826 LyGlnGlnGlyValThrGlnValaLeuLysGlyAlaProGlyGlnPro 842

```

1145 .....TGGTAGTACCAATGCGACACATGCGCGATGCGCGAC 1108  
843 G1YthrlleuValGthValrPrometle1yGla1dgrleuValThrPr 859  
1107 GTTGACG.....GCGTGTGTAACTGTGAAGATTGT 1076  
859 OvalThrValSerAlaValLysProAlaValThrThleu..... 872  
1075 TTTTCAGAAATGGCGCGCGGTGCTACCG..... 1047  
873 .....ValYallYsg1yThrThrG1yValThrThr 882  
1046 .....GTGATGAGATATTGTTCGCGCGCGCG 1018  
883 LeuG1yThrValThrG1yThrValSerThrSerLeuAla1yG1 899  
1017 AACCGACCGCAACGCTTTTTCGCGCGCGCTTCTGTGATACGAAATCT 968  
899 YH1SerThrSerAlaSerleu..... 906  
967 GATTGTGTAGCTCCCAATAATCGTGGCGCTTGTGAATCGCGCG 918  
907 .....AlaThrPr 909  
917 TTGATATCGCAACGGAAATACGCGGTGTGTGCGTACCAATTCGCC 868  
910 I1etThrThrLeuG1yThrThrleAlaThrLeuSerSeG1nAl11eAsPr 926  
867 CGCAGTAATTTGCGATACCTTTGCGACCCAAACGGGTACGCAAGACGCTG 818  
926 oThrAla1leThrValSerAla1aG1ThrThrleuThrAla1a1yG 943  
817 GTTGG.....TTCAGTTAGACCA.....CCCAACGA 789  
943 1yLeuThrThrProThrThrleThrMetG1nProValSerG1nProThrG1n 959  
788 ATGACG.....CGCTGGGTTCACACGCGCTGTT.... 759  
960 ValThrleu1leThrAlaProSerG1yAla1u1aG1nProVal1n1as 976  
758 .....GCAACAAACGTCGATGGCA 737  
976 PleuProValSer1leLeuAlaSerProThrThrG1nG1nProThrAlaT 993  
736 TTACATCTGATTAATGATGCTCCAAACGCT..... 706  
993 hr.ValThr1leAlaSpSerG1yG1nG1aSerValG1nProG1yThru 1009  
705 .....TTTGTTCGACCGAC.....CGGTCATCAATGATGTTG 670  
1009 alThrleuValCysSerAsnProProCysG1uThrThn1sG1uThrG1yThr 1025  
669 CGTGCACATCAACCGCGCGGATGCGCGG..... 640  
1026 ThrasThrAlaThrThrThrThrValAla1aAsnLeuG1yH1sProG1 1042  
639 .....GCCGAATCATGTGT..... 625  
1042 nProThrG1nValG1nPhrValCysAsp1rG1nG1uAla1a1a1aSer1 1059  
624 .....TTCGATGTTGGACAGATTTTCGACGCGCAC 595  
1059 euValThrSerThrValG1yG1nG1nAsnG1ySerVala1aTrg..... 1073  
594 GTCTGGCGCAC.....TGCTTACACACATGATTTTACG..... 559  
1074 ValCysSerAsnProProCysG1uThrThn1sG1uThrG1yThrThasnth 1090  
558 .....CTCGGTCA..... 550  
1090 rAlaThrThrAlaThrSerAsnMetAlaG1yG1nH1sG1yCysSerAsnPr 1107



```

COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/393,703
FILING DATE: 24-FEB-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A.
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57503-2/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-393-703-5

alignment_scores:
Quality: 100.50 Length: 634
Ratio: 0.393 Gaps: 32
Percent Similarity: 40.379 Percent Identity: 20.662

alignment_block:
US-09-303-518D-127/rev x US-08-393-703-5 ..

Align seg 1/1 to: US-08-393-703-5 from: 1 to: 2035

1343 CACGCTTCCTTCGCANAGGTTTCACGACCTTACGCAACAGCGGCCNTA 1294
747 LysProThrIleuGlyIleSerSerVal.....SerProse 759
1293 TTGSTATTTGCCGCGACAGCAGAGCTGCACAAAGCGAGTCTTTCGT 1244
::: ||||| ||::: ||||::: ||:
759 rthrThrLysProGlyThrThrThrIleIleLysThrIleProMetSerA 776
1243 CCATTCGCAACGCAACCCAAATGCTGGCGGCTGCGGTATGCGGACGATT 1194
::: ||||::: ||||::: ||||| |||
776 laIleIlethrGlnAlaGlyAlaThrGlyValThrSerSerProGlyIle 792
::: ||||::: ||||::: ||||| |||
1193 AAATCG..... 1188
|||||
793 LysSerProIleThrIleIleThrThrLysValMetThrSerGlyThrG1 809
1187 ..... CCGCAAGACGAGGTAGCGAGA 1166
|||
809 yAlaProAlaLysIleIleIleThrAlaValProLysIleAlaThrIleYlsg 826
1165 TGCTAGGCGGATTCAGCGC..... 1146
::: |||||::: ||||:::
826 LysGlnGlnGlyValThrGlnValValLeuLysGlyAlaProGlyGlnPro 842
1145 ..... TCGTAAGTACCAATGCGGCAACGATGGCGGCTGGCCACC 1108
|||||::: ||||::: |||
843 GlyThrIleLeuArgThrValProMetGlyValAlaArgLeuValThrPr 859
1107 GTTAGCG..... GCGTGCGTGAAGTGTGT 1076
||::: |||||::: |||||
859 oValThrValSerAlaValAlaLysProAlaValThrThrLeu..... 872
1075 TTTTCAGCAAAATGCGCGAGGCTCGTACGC..... 1047
|||||:::

```

```

873 .....ValValLysGlyThrThrGlyValThrThr 882
1046 .....GTGATGAGTATTTGTCGGCTGCGGCGC 1018
883 LeuGlyThrValThrGlyThrValSerThrSerLeuAlaGlyAlaGly 899
1017 AACCCGACGACAGCTCTTGGCTCGGCGCTTTCGATACGGAATCT 968
899 yHisSerThrSerAlaSerLeu..... 906
967 GATGTGTAGCGTCCAAATATATGTCGCGCTGTGTATGCGGCGC 918
907 .....AlaThrPro 909
917 TTGATACCGAACGGAATACCGGTTGTCTGCGTCAACCAATTCGCC 868
910 IleThrThrLeuGlyThrIleAlaThrLeuSerSerGlnValIleAsnPr 926
867 CGCAGTAAATTTGCGATNCTTCGACCCAAACGATGCGCAAGCGCGT 818
926 oThrAlaIleThrValSerAlaAlaGlnThrThrLeuThrAlaAlaGly 943
817 GTTTG.....TTGACTTGAGACCA.....GCCAAGCA 789
943 LysLeuThrThrProThrIleThrMetGlnProValSerGlnProThrGln 959
788 ATCAG.....CGCTCGTGTTCAGACGCGCTGT..... 759
960 ValThrLeuIleThrAlaProSerGlyValGlnAlaGlnProValHis 976
758 .....GCAACCAACGTCGAGCGCA 737
976 PleuProValSerIleLeuAlaSerProThrThrGlnGlnProThrAla 993
736 TTACATCTGATATGATGATGTCGCAACGCT..... 706
993 hr...ValThrIleAlaAspSerGlyGlnGlyAspValGlnProGlyThr 1009
705 .....TTGTGTGACCGAC.....CGGCTCAATGAAATGATGTG 670
1009 alThrLeuValLysSerAsnProProCysGluThrHisGluThrGlyThr 1025
669 CGTGCACCTCAACCGCGCGGATGCGGCGC..... 640
1026 ThrAsnThrAlaThrThrThrValAlaAlaAsnLeuGlyGlyHisProG 1042
639 .....GCCGAATCATGTGT..... 625
1042 nProThrGlnValGlnPheValCysAspArgGlnGlnAlaAlaSerL 1059
624 .....TTGATGTTGGCAGCATTTTCAGACGGCAC 595
1059 euValThrSerThrValGlyGlnGlnAsnGlySerValAlaArg..... 1073
594 GTCTGCGCGCAGC.....TGCCTACACACATGATTTTAG..... 559
1074 ValCysSerAsnProProCysGluThrHisGluThrGlyThrHisAsnTh 1090
558 .....CTCGGTCAA..... 550
1090 rAlaThrThrAlaThrSerAsnMetAlaGlyGlnHisLysSerAsnPr 1107
549 .....ACGGCTCAATACACAGACACACATTCG 526
1107 roProCysGluThrHisGluThrGlyThrThrAsnThrAlaThrThrAl 1123
525 TCTGAATCMTGCGCGCT..... 507
1123 ameterSerValGlyAlaAsnHisGlnArgAspAlaArgAlaCysA 1140
506 .....TCTTGATCACACACACACGCGG 486
1140 laAlaGlyThrProAlaValIleArgLysSerValAlaThrGlyAlaLeu 1156
485 TCTGCCGACGACGCA.....TT 469
1157 GlnAlaAlaGlnGlySerLysSerGlnCysGlnThrArgGlnThrSerAl 1173
468 GGTGTCCATTCGATTCAGAGATGGCGAAGCGGTGCGATCGACGGCG 419
1173 aThrSerThrThrMetThrValMetAlaThrGlyAlaProCysSerAla 1190
418 GGAATTTGCTG.....AACGACCGATTCACG 393
1190 LysProLeuLeuGlyProSerMetAlaArgLysProGlyGlyArg..... 1204
392 AGCGAGTCCCAACCGGATTCGATTCAGATTCGTCNNCANNATTCNTGCC 343
1205 .....SerProAlaPheValGlnLeuAlaProLeuSerSerLys 1217
342 GCTTAAGTTTGGCCAAACGCTTCGGGCGGTGCTTGGACCTGATTCGT 293
1217 sValArgLeuSer.....SerProSerIleLysA 1227
292 CGTTGCTT.....TCAACGCAATTCAGACGCGATGAGTACG 255
1227 spleuProAlaGlyArgHisSerHisAlaValSerThrAlaAlaMetThr 1243
254 CGCTTTTCG.....CCGCGATGGATGCGCGGATTTTGGC 220
1244 ArgSerSerValGlyAlaGlyGlnProArg...MetAlaProValCysG 1259
219 T.....GAACGCGCGGTAAACACGACCGCGGATCTTTTGTCTT 176
1259 userLeuGlnGlyLysProSerThrThr.....ValThr 1272
175 CAACACGACCTTGGCCCTTTTTCAGCGCATCGCTTCCTGACTTCATC 126
1272 alThrAlaLeuGlnAlaLeuLeuCysProSerAlaThrValThrGlnVal 1288
125 NAGGCGCGCATACCGCATATCTTCGCCACGACGACCTCGGTAAT 76
1289 CysSerAsnProProCysGluThrHisGluThrGlyThrThrAsnThrAl 1305
75 GACG.....GGCCGCTCATTAATGACTTCGCTGCTGCC 39
1305 aThrThrSerAsnAlaGlySerAlaGlnArgValCysSerAsnProPro 1321
seq_name: /cgn2_6/p/oddata/1/iaa/PCTUS_COMB.pep:PCT-US93-11721-5
seq_documentation_block:
: Sequence 5, Application PC/TUS9311721
: GENERAL INFORMATION:
: APPLICANT: Lamarco, Kelly
: APPLICANT: Wilson, Angus
: APPLICANT: Heir, Winship
: TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
: TITLE OF INVENTION: HOST CELL FACTOR
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: FLEHR, HONBACH, TEST, ALBRITTON & HERBERT
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/11721
: FILING DATE: 03-DEC-1993
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:

```



392 AGCGACAGTCCACAACCGGATTGATTCAGATTGCNNNNNANTTCGCCC 343  
1205 .....SerProAlaPheValGlnLeuAlaProLeuSerSerLy 1217  
342 GCTTAAGTTTGCCACACGCTGGCGGCGCTAGCGTTTCGACGATTTCGT 293  
1217 sValArgLeuSer.....SerProSerIleLySA 1227  
292 CGTTGCTT.....TCACGGCATACACGCGACTGAGTAAGC 255  
1227 spleuProAlaGlyArgHisSerHisAlaValSerThrAlaAlaMetThr 1243  
254 CGCTTTTCG.....CCGCGATGATGGCGGCGATTTCGCC 220  
1244 ArgSerSerValGlyAlaGlyLupProArg...MetaLapProValCysG 1259  
219 T.....GAACNGCGCGGTAACACACCGCGCGATTCCTTTGCTT 176  
1259 uSerLeuGlnGlySerProSerThrThr.....ValThrV 1272  
175 CAACACAGCAGCTGGCTTTTTCGAGCGCATCGCTTCCTGACTTCATC 126  
1272 alThrAlaLeuGlnAlaLeuLeuCysProSerAlaThrValThrGlnVal 1288  
125 NAGGCGCGCATACCGCATATTCCTGCCACAGACGCGACTTCGTAAT 76  
1289 CysSerSerInProProCysGlnThrHisGlnThrGlyThrThrAsnThrAl 1305  
75 GAGC.....GGCCGCTCATPAAATGACTTCGCTCGCTGCTGCC 39  
1305 alThrThrSerAsnAlaGlySerAlaGlnArgValCysSerAsnProPro 1321  
seq\_name: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:US-08-270-076A-11  
seq\_documentation\_block:  
Sequence 11, Application US/08270076A  
Patent No. 5667986  
GENERAL INFORMATION:  
APPLICANT: Sleep, Darrell  
APPLICANT: Goodey, Andrew R  
APPLICANT: Vakeria, Diana  
TITLE OF INVENTION: Yeast Promoter  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: The BOC Group, Inc.  
STREET: 100 Mountain Avenue, Murray Hill  
CITY: New Providence  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07974  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/270,076A  
FILING DATE: 01-JUL-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 8933521.2  
FILING DATE: 18-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/597,687  
FILING DATE: 16-OCT-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/925,286  
FILING DATE: 04-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Swope, R. Halin  
REGISTRATION NUMBER: 24864  
REFERENCE/DOCKET NUMBER: 92H834-3

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 908/771-6292  
TELEFAX: 908/771-6159  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 806 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-270-076A-11  
alignment\_scores:  
Quality: 99.00 Length: 407  
Ratio: 0.556 Gaps: 17  
Percent Similarity: 43.735 Percent Identity: 22.113  
alignment\_block:  
US-09-303-518d-127/rev x US-08-270-076A-11 ..  
Align seg 1/1 to: US-08-270-076A-11 from: 1 to: 806  
1322 TCACGACCTTACGCAAGCGCGCCNTATTGTAATTGCCGCGCAGAC 1273  
|||||.....:||||| |||  
33 SerSerAsnIleThrSerSerClyProSerSer.....Th 44  
1272 GAAGTGCACAACGAGCTCTT.....T 1247  
44 PProPheSerSerAlaThrGlnSerPheSerThrGlyThrThrValThrP 61  
1246 CGTCCAAATTCACAGCAACCAATGCT...TCGCGCGTGGTATTCGCC 1200  
|||||.....:||||| |||  
61 roSerSerSerLySerProClySerLyThrGlnThrSerValSerSer 77  
1199 ACGATTAATCGCCCAAGCAGGTGTTTTCAGAAATGCGGAGGTCGTA 1150  
|||.....: ||| ||| |||  
78 ThrThrGlnThrThrIleValProThrThrThrThrSerValIleTh 94  
1149 GCGCTCGTAAGTACCAATCGGCACCATGGCGCGCTCGCCACCGTACGG 1100  
||| ..: ||| ||| ..: |||  
94 rProSerThrThrThrIleThrThrValCysSerThrGlyThrAsn 111  
1099 CTGTCGTGAAGCTTGAGAGTTGTTTTCAGAAATGCGGAGGTCGTA 1050  
: ..:  
111 er..... 111  
1049 CGCGTATGAGTATTTCGCGGCTGGGCGCAACCGCAAGCTC 1000  
: ..: ||| ||| ||| |||  
112 ...AlaGlyLupThrThrSerGlyCysSer..... 120  
999 TTGCTCGCGCTTCGATACGGAATCGATGTGTGTAGCTGCCA 950  
|||||.....: ||| ||| |||  
121 .....ProLySerThrIleThr..... 125  
949 AATAATCGTCGCGCTGTGTAATGCGCGCTCAATACGACCGAA 900  
: ..: ||| ||| ..: ||| |||  
126 ....ThrThrValProCysSerThrSerProSerLupThrAlaSerGlu 140  
899 ATCAGCGGTGCTGCTGCAACCAATTCGCCGCGCATATTTCGATAC 850  
||| ||| ..: ||| |||  
141 SerThrThrThrSerProThrThrProValThrThrValAlaThrThr 157  
849 TTTCGCAACC.....AAACGGTACCAAGAGCGCTGCT..... 816  
: ..: ||| ||| ..: ||| |||  
157 rValValThrThrGlyThrSerThrSerThrLysGlnGlyGluIleT 174  
815 .....TTGTTGACTTGA 804  
174 hrThrThrPheValThrLysAsnSerProThrThrThrLeuThrThrIle 190  
803 GAACCAACCAAGCAATCAGCGCTGCTGTTCAACGCGCTGTGCAAA 754  
||| .....: ||| ||| ||| |||  
191 AlaProThrSerSerValThr.....ThrValThrAs 201

```

753 CAACGCTCGATGSCATATCATCTTGATATATGATGTCACAAAGCTTT 704
    |||  ....
201 nphethrprothrthrlthrthr.....ThValc 212
703 TGTTCGACGCGACCGCTCAATGAAATGAGTGGCTGCACCTCAACCG 654
    :|||:
212 ysserthnlythrAnsSer.....AlaGlyIuThrThr 223
653 GCCGAGTCGCGCGCGCAATTCATGTGTTGATGTTGAGTGGACATTTTC 604
    :|||:
224 SerGlyCysSerProlyThrValThrThrThValLeuYsserThrGl 240
603 AGACGGC.....ACGTGCGCGCAGCTGCTTACACACATGATTTT 563
    :|||:
240 YthrGlyIuThrThrThrThrGlAlaThrAlaProValThrThAlaVal 257
562 TACGCTCGGTCAACGCGCTCAATACACGACANACNTGCTGAAATCTTCG 513
    :|||:
257 hrThrThrVal..... 260
512 NCGGCTTCTTGATGACCAACGAGGTCTGCGCGNAGCGGATTTGTTTC 463
    :|||:
261 .....ValThrThrGlUsserThrThrThrAnsSerValGl 273
462 CATCGCATTTGACGAGATGGCGAACGCTCGCATTCAGAGCA..... 420
    :|||:
273 YlystrThrThrThrSerYThrThrThrLysSerValProThrThrYValP 290
419 .....GGGATTTTG.....CTGAC 405
290 heAspPheGlyLysGlyLeuAspGlnSerCysGlyGlyValPheSer 306
404 GGACGAGNTAGCGACGCGCTCCACAAACGCGATTTGATGAGATTTGNCN 355
    :|||:
307 AsnAsnGlySerSerGlnValGlnLeuArgAspValValLeuMet..... 321
354 NANTTCGCGCGCTTAAGTTTGCACACGCTTCGGCGCGCTAGGCTTGA 305
    :|||:
322 .....AsnGlyThrValValYlyrAspSerAsnGlyAlaTrpAspSer 336
304 ACTCGATTTGCTGCTGCTTCAGCGCAATCAACGCGACGACTGAGTACG 255
    :|||:
336 eAlaLeu.GluGluTrpLeuGlnArgGlnYlyLysValSerIleGluAr 352
254 CGCTTTCGCGCGGATGGA 236
    :|||:
352 glIlePheGluAsnIleGly 358
seq_name: /cgn2_6/ptodata/1/1aa/5B-COMB.pep:US-08-760-489-2
seq_documentation_block:
; Sequence 2, Application US/08760489
; Patent No. 5830696
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC
; TITLE OF INVENTION: ENZYMES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/760,489

```

```

FILING DATE: 05-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/008,311
FILING DATE: 07-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Hallie, Ph.D., Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/008001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1487 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-760-489-2

alignment_scores:
Quality: 97.50 Length: 440
Ratio: 0.516 Gaps: 23
Percent Similarity: 42.955 Percent Identity: 23.409

alignment_block:
US-09-303-518d-127 x US-08-760-489-2 ..

Align seg 1/1 to: US-08-760-489-2 from: 1 to: 1487

219 AGGCAAAATCGCCGCAT..... 236
|||
355 ArgGlnAsnArgArgHisAlaArgAlaArgArgAspSerGluSer 371
237 .....CCATCGCGCGGAAAGCGCGTCTTCAGTCG 270
|||
371 rGlyAspSerLeuLeuProProArgTrpArgAlaProGlnIleAlaGly 388
271 GTGATTCGCGTTGAGAGCAGCAGCAAAATCGAGTTGCAACGCTACGCGC 320
|||
388 euGlnIleHisArgGlnProGlnProLeuHisAlaProProLeuAlaSer 404
|||
321 CGAAGC.....GTGG 331
|||
405 ArgSerAspAsnArgGluGlyGluGlyProHisGlnArgAspTrpLeuGl 421
332 CAACCTTAAGCGCGCGANGAANTNGNNGCATCTGATCCATCGCGTTG 381
|||
421 yArgAlaGlyGluArgGlyGlyValArgGln.....ValArg.... 433
|||
382 TGGACTCGCTGCGGTANCCGCTCCGTTACAGCAAAATCCGTCGTCGATGC 431
|||
434 .....GlyValHisArgGlnArgThrArgGlyPro 443
432 CGACCGCTTGCATCTTCTGTCATATGCGATGACACCAATCCGCTNCGCG 481
|||
444 ArgTrpValGlnHisLeuGlnArgAlaAspGlyArg..... 455
482 CAGACCGCTGGTGTGTGATCAAGAAGCCGNCANAGATTTTCAGACGANGT 531
|||
456 .....CysGlyVala...ArgLeuProArgAlaLeuLeuArgLeuSerA 469
532 NTGCTGTATTGAGCGG...TTTGACCGAGCGTAATAATCATGTGTGTGA 578
|||
469 IaglyGlyTrpGluProArgGlyGlyLysAlaGlyAsnPro..... 482
579 GCGAGCTGGCGGACAGACGTGCGGTGAATAATGCTGCACATCGAACAAC 628
|||
483 .....GlnHisAsp..... 485
629 ATGATTCGCGCGCGCGCATCC..... 650

```

```

486 .....:|||||:
      lysargproargthnglyleuglnaspraplsylvalarg 500
651 .....:|||||:
      lnglylsylvalarglypserleuargglyargglyargaspasyl 516
500 lnglylsylvalarglypserleuargglyargglyargaspasyl 516
666 CACGACATTCATTCATTGAGCGCGCGGTGCAACAAACCGTTGGA 715
      :|||:
517 lnglnlnlsarg.....Argcysleuserllearglyleugl 528
716 CCATCAATTATCAGATGTATGTC.....CATC 744
      :|||:
528 nagrprolyslylvalarglysercysarglylvalarglnleuarglna 545
745 GCACGTTGTTGTCACAGCGCG.....TCGACACCGCGCGCGGT 785
      :|||:
545 tgalaleuleuargargasnproglnglylnlaaglnlnlsargvalarg 561
786 GATGCTTTGGGTGCTTCAGTCACAAACAAACACCGCTTGGCTGACG 835
      :|||:
562 Argasprleuarglnserleuarglnlylvalargleuargprarg 578
836 T.....TTGGGTGCGAA..... 848
578 glnleuarglnvalargserarglnvalarglnvalarglnvalargln 595
849 .....AGATCGCAATTTACTGCGCGCAATTTGTTGACGACAGCA 890
595 lshproasprlleuarglnserarglnleuarglnleuargln 611
891 CCGCGGATTTCCGTTGTTGAAACCGCGATTTACACAGCGCGCG 940
      :|||:
612 Alarglnphe.....Pheargarglylsalalarglylsargln 625
941 ACATATTGTTGACGCGCTACCAACATCAGATTTCCGTTATGCAAGAGC 990
625 shargleu...glyasprleuargglylvalargleuarglylsarg 641
991 CCGACCAAGACGCTGTTGCGTGGTGTG.....CGCGCGACCG 1028
      :|||:
641 lnglnlnlnlearg..glyproglyleuarglnlsarglylsarglnsarg 657
1029 GCACAAATTCATTCACGCGCTACGACCGCTGCG.....CCATTCCTGAAA 1075
      :|||:
657 tgrpshlnlsproalalalvalleu..proarglnprocyssergluas 673
1076 ACAAACTCTCTACGACGACGCGCTACGCGCGCGCGCGCGCATG 1125
      :|||:
673 pelylvalarg..glylyleuargarglnlyleuargproleuarg 689
1126 GTCCGATTTGTTGACTTACGACGCGCTAATGCGGTACGATCTGCGCTAC 1175
      :|||:
689 lvalaasprarglnleuargvalarglyproarglnlsargglyvalarg 705
1176 CCGTGTGTTGCGGATTTAATCGTGGCGGATACCGACGCGCGCAAGCAT 1225
      :|||:
706 proleuserglyserlnlsasnnglnly.....gluasnhrhalaglyl 720
1226 TGGGTGTTGCGGATTTGACGACGAGCAACGCTGTTGTCAGCTCGTC 1275
      :|||:
720 ylsarglylsylleuarglnlsargglylnlaargserglnglyl 737
1276 TGCCCGGCGCAATTCGA 1292
      :|||:
737 snproglylvalarg 742

```

seq\_name: /cgn2\_6/ptodata/1/aa/5B\_COMB.pep:US-08-760-489-4

seq\_documentation\_block:

; Sequence 4, Application US/08760489  
; Patent No. 5830696  
; GENERAL INFORMATION:

```

? APPLICANT: Short, Jay M.
? TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC
? TITLE OF INVENTION: ENZYMES
? NUMBER OF SEQUENCES: 4
? CORRESPONDENCE ADDRESS:
? ADDRESS: Fish & Richardson, P.C.
? STREET: 4225 Executive Square, Suite 1400
? CITY: La Jolla
? STATE: CA
? COUNTRY: US
? ZIP: 92037
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: Windows95
? SOFTWARE: FASTSEQ for Windows Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/760,489
? FILING DATE: 05-DEC-1996
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 60/008,311
? FILING DATE: 07-DEC-1995
? ATTORNEY/AGENT INFORMATION:
? NAME: Haile, Ph.D., Lisa A.
? REGISTRATION NUMBER: 38,347
? REFERENCE/DOCKET NUMBER: 09010/008001
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 619/678-5070
? TELEFAX: 619/678-5099
? INFORMATION FOR SEQ ID NO: 4:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1487 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? FRAGMENT TYPE: internal
US-08-760-489-4

```

alignment\_scores: Quality: 97.50 Length: 440  
Ratio: 0.516 Gaps: 23  
Percent Similarity: 42.955 Percent Identity: 23.409

alignment\_block:  
US-09-303-518d-127 x US-08-760-489-4 ..

Align seg 1/1 to: US-08-760-489-4 from: 1 to: 1487

```

219 AGCGAAATTCGCGCGCAT..... 236
      :|||:
355 Arglnlnsnargarglnsalalargalargalargalargaspsergluser 371
237 .....:|||||:
      :|||:
371 rglaspserleuargproargargargargargargargargargarg 388
271 GTGATTCGCTGTAAGGCAACGCAAAATCGAGTTGCAACGCTACGCGCG 320
388 euglnnglylnlsargglylvalargln.....Valarg.... 433
321 CGAACG.....GTGG 331
      :|||:
405 Argserasprasnarglylvalarglylvalargprohlsignatargsp 421
332 CAACCTTAAGCGCGCGGANGAANTNNGNCATGTGATCGAATCGCGTTTG 381
      :|||:
421 yargalaglylvalargglylvalargln.....Valarg.... 433
382 TGGACTGCGCTCGTANCGCGTCCGTCAGCAAAATCCCTGCGCTGATGC 431
434 .....:|||||:
      :|||:

```

```

432 CGAGCCGTTCCGCAATCTTGCATTCGATGACACCAATCCGCTGCGG 481
444 ATGTGTATGAGTATGATCAAGAGCCGCGATTTTCAAGAGT 455
482 CAGACCCGTTGTTGATCAAGAGCCGCGATTTTCAAGAGT 531
456 .....CysGlyAla...ArgLeuProArgAlaLeuLeuArgLeuSer 469
532 TTGACCGAGCGTAAATCCATGCTGTGTAA 578
469 TactGlyGlyArgProArgGlyGlyAlaGlyAspPro..... 482
579 GGCAGCTGGCGCAGAGTCGCGTGAATAATGTCACATCGAAGAC 628
483 .....GlnHisAsp..... 485
629 ATGAATTCGGCGCGCCATTC..... 650
486 .....LysArgProArgThrGlyLeuGlnAspArgGlyValArg 500
651 .....GGCCGTTTGAGTGG 665
500 InGlyGlyArgGlyPheProLeuArgGlyArgGlyArgAspAsnLeu 516
666 CAGCAGCATTCATTCATTGAGCGGTCGCGTCAAAACAAACGTTTGA 715
517 GlnGlnHisArg.....ArgCysLeuSerIleArgLeuGln 528
716 CCATCAATATTCAGATGTAATTC.....CANC 744
528 nArgProArgGlyArgGlyLeuSerCysArgGlyArgGlnLeuProGln 545
745 GAGCTGTTGTTGCAAGAGCGG.....TCTGAACCGCGAGCGCT 785
545 rglAlaLeuLeuArgArgAsnProGlnGlyGlnAlaGlnHisArgValArg 561
786 GATTGCTTGGCGTTCCTCAAGTCAACAAACGCGCTTGGCGTACG 835
562 ArgAspLeuArgGlnSerSerAlaSerGlnGlyGlnArgLeuAspArg 578
836 T.....TTTGGTGGCGAA..... 848
578 GLeuLeuHisGlyArgSerArgGlnValPheGlyAlaGlnValProGln 595
849 .....AGATCCGCAATTTACTGCGGCGAATTTGTTGACGCAAGAC 890
595 IsthProAspIleLeuProGlySerSerGlnLeuArgLeuArgGln 611
891 CCGCGTGAATTCGCGTTCGATTAAGAGCGCGGATTAACAAAGCGCGC 940
612 AlaArgGlnPhe.....PheArgArgGlyLysAlaArgLysArg 625
941 AGCATTTATTTGGAGCTACACCAATCAAGATTCGTTATGAGAGAGC 990
625 sArgLeu...GlyAspLeuSerGlyGlyAspLeuArgLeuAspLysArg 641
991 CCGAGCAAGAGCTGTCGCTGGGTG.....CGCGCAGCGC 1028
641 LysGlnGlnIleArg...GlyProGlyLeuArgHisArgLysArgAsnSer 657
1029 GAGCAATATTCGATTCAGCGTACGAGCCCTCGG...CCATTTCTGAAA 1075
657 rghPheAsnHisProAlaValLeu...ProArgGlnProCysSerGlnAs 673
1076 ACAAACTCTTCAGATTCAGCAGACGCGTCAACGCGTGGAGCGCGCAT 1125
673 rglGlyValArg...GlyLeuLeuArgGlnGlyLeuProLeuLeuG 689
1126 GTGCGGATTTGATCTTACGAGCGCGTAAATCCGCTAGACATCTCTCC 1175
689 LysAlaAspArgGlnLeuArgValGlyProArgPheGlnAspGlnValArg 705

```

```

1176 CCGCTTTTGGCGATTTATGCTCGGATACGACAGCGCGCAAGCAT 1225
706 ProLeuSerGlySerHisAsnGlnGly.....GlnAsnThrAlaGly 720
1226 TGGCTGCTTGAATTCGAGAGAGAGCTTCGCTTGTGCAGCTTCGTC 1275
720 LysArgGlyGlyLeuGlnHisArgGlyGlnGlnArgSerGlnGlnGly 737
1276 TGCCCGGCAATACGA 1292
737 snProGlyGlnValArg 742

```

```

seq_name: /cgn2_6/ptodata/1/aa/6E_COMB.pep:US-09-185-373-2

```

```

seq_documentation_block:
: Sequence 2, Application US/09185373
: Patent No. 6335179

```

```

GENERAL INFORMATION:
APPLICANT: Short, Jay M.
TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC
ENZYMES

```

```

NUMBER OF SEQUENCES: 4

```

```

CORRESPONDENCE ADDRESS:
ADDRESSER: Fish & Richardson, P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: US
ZIP: 92037

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/185,373
FILING DATE: 03-No. 6335179-1998
CLASSIFICATION: <Unknown>

```

```

PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/08/760,489
FILING DATE: 05-DEC-1996
APPLICATION NUMBER: 60/008,311
FILING DATE: 07-DEC-1995

```

```

ATTORNEY/AGENT INFORMATION:
NAME: Hallie, Ph.D., Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/008001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099

```

```

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1487 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```

```

US-09-185-373-2

```

```

alignment_scores:
Quality: 97.50 Length: 440
Ratio: 0.516 Gaps: 23
Percent Similarity: 42.955 Percent Identity: 23.409

```

```

alignment_block:
US-09-303-518d-127 x US-09-185-373-2 ..
Align seg 1/1 to: US-09-185-373-2 from: 1 to: 1487

```

```

219 AGGCAAAATTCGCGCAT..... 236

```



```

355 ArgGlnAsnArgHisAlaIleArgAlaIleArgAspSerGlnSer 371
237 .....CCATGCGCGCGGAAAGCGCGTACTTCACTGCGTC 270
371 rGlyAspSerLeuLeuProProlArgIleArgAlaProlGlnGly 388
271 GTGATGCGCGTTGAAGCGCAAGAAATCGAGTTCGACGCGCGCGC 320
388 euGlnGlyHisArgGlnProlProlProlHisAlaProlProlAsp 404
321 CGAAGC.....GTTGG 331
405 ArgSerAspAsnArgGlnGlyGlyProHisGlnArgAspTrpLeuG 421
332 CAACCTTAAGCGCGGANGAANTNNGNCAATCTGATCCATCGCGTTG 381
421 yArgAlaGlyIleArgGlyGlyValArgGln.....ValArg.... 433
382 TGCACTGCGCGTANCCGTCGTTGACAAATCCCTGCGCGTACATGC 431
434 .....GlyValHisArgGlnArgTrpArgGlyPro 443
432 CGAGCGGTCGCGCATCTTCTGCAATCGATGACACCAATCGCTNCGG 481
444 ArgTrpValGlnHisLeuGlnArgAlaAspGlyArg..... 455
482 CAGACCGCTGCGTGTGATCAAGAAGCGCGANGATTCAGACGANGT 531
456 .....CysGlyAla...ArgLeuProArgAlaLeuLeuArgLeuSer 469
532 NTGCGTGTGATGAGCGG...TTTGACCGAGGTAATCCATCTGTGTGA 578
469 IAGIlyGlyTrpGlnProlArgGlyGlyIleValAsnPro..... 482
579 GGCACGTCGCGCAGACGTCGCGTGTGAAATGTCGCCAATCGAAGAC 628
483 .....GlnHisAsp..... 485
629 ATGAATCGCGCGCGCCGATCC..... 650
486 .....LysArgProArgTrpGlnLeuGlnAspAspGlyValArg 500
651 .....GCCGCGTTGAGTGG 665
500 lngIlyGlyArgGlyPheProlLeuArgGlyArgGlyArgAspAsnLeu 516
666 CAGGACATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 715
517 GlnGlnHisArg.....ArgCysLeuSerIleArgGlnLeuG 528
716 CCATCAATTATCAAGATGTAATTC.....CANC 744
528 nArgProlGlyArgGlnSerCysArgGlyArgGlnLeuLeuProlGln 545
745 GGACGTTGTTTGCACAGCGCG.....TCTGAACACGCGCGCGT 785
545 rGAlleuLeuArgArgAsnProGlnGlnGlnHisArgValArg 561
786 GATTCGTTTGGTGTCTTCAGTCAACACACACCGCTTGCAGTACCG 835
562 ArgAspLeuArgGlnSerSerAlaSerGlnGlyGlnArgLeuAspArg 578
836 T.....TTTGGTGCAGAA..... 848
578 gLeuLeuHisGlnArgSerArgGlnValPheGlyAlaGlnValProlGln 595
849 .....AGTATCGCAAAATTACTGCGCGCAATGCTTGAACGACAA 890
595 IsthProlAspIleLeuProGlySerSerGlnLeuArgLeuArgGln 611
891 CCGCGGATTCGCGTTCGATTAAGCGCGCGATTCACACAGCGCGCG 940
612 AlaArgGlnPhe.....PheArgArgArgGlyAlaArgGlyAspArg 625

```

```

941 ACGATTATTTGGAGCGCTTACCAATACAGATTTCCTTATCGAAGAGC 990
625 sArgLeu...GlyAspLeuSerGlyGlyAspLeuArgLeuAspArg 641
991 CGACGCAAGAGCTGTGCGTGGTGG.....CGCCGACGCC 1028
641 yGlnGlnLeuArg...GlyProlGlyLeuArgHisArgGlyAspAsnSer 657
1029 GGACAAATATTCATTCACCGCTTACGACCGCTGG...CCATTTCTGAAGA 1075
657 rGPhAsnHisProlAlaValLeu...ProlArgGlnProlCysSerGln 673
1076 ACAAACTTTCAGTTCACGACAGCGCTCAAGCGCGCGCGCGCGCATG 1125
673 rGlyGlyValArg...GlyGlyLeuArgArgGlnGlnLeuProlLeuG 689
1126 GTGCCGATGCTACTTACGAGCGCGTAAATGCGCTAGACATTCCTTAC 1175
689 yAlaAspArgGlnLeuArgValGlyProlArgPheGlnAspGlyValArg 705
1176 CCTCTTTTCCGCGATTATTCGTCGCGATACGACAGCGCGGACGAT 1225
706 ProlLeuSerGlySerHisAsnGlnGly.....GlnAsnTrpAlaGly 720
1226 TGCGTTGCTTGAATGAGCAAGAAGACCTCGCTTGTGCGAGCTTGTG 1275
720 yLysArgGlyGlyLeuGlnHisArgGlyGlnGlnArgSerGlnGlnGly 737
1276 TGCCGCGCGCAATACGA 1292
737 snProlGlyValArg 742

seq_name: /cgn2_6/prodata/1/iaa/6B_COMB.pep:US-09-185-373-4
seq_documentation block:
; Sequence 4, Application US/09185373
; Patent No. 635179
GENERAL INFORMATION:
APPLICANT: Short, Jay M.
TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC
ENZYMES
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: Fish & Richardson, P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: US
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTED for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/185,373
FILING DATE: 03-NO-635179-1998
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/760,489
FILING DATE: 05-DEC-1996
APPLICATION NUMBER: 60/008,311
FILING DATE: 07-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Hallie, Ph.D. Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/008001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:

```

```

?      LENGTH: 1487 amino acids
?      TYPE: amino acid
?      TOPOLOGY: linear
?      MOLECULE TYPE: protein
?      FRAGMENT TYPE: internal
?      SEQUENCE DESCRIPTION: SEQ ID NO: 4
US-09-185-373-4

```

alignment_scores:	
Quality:	97.50
Ratio:	0.516
Percent Similarity:	42.955
	Length: 440
	Caps: 23
	Percent Identity: 23.409

```
alignment_block:
US-09-303-518D-127 x US-09-185-373-4
```

Align seg 1/1 to: US-09-185-373-4 from: 1 to: 1487

```

219 AGCGAAATTCGCCGAT..... 236
      |||||
355 ArgGlnAsnArgArgHisAlaIleValArgAlaAspSerGluSerSe 371
237 .....CCATCGCGGGCAAAAGCCGCTACTTCAGTCGGTC 270
      ||| |||
371 rGlyAspSerLeuLeuProArgTrpArgTrpAlaProGlnIleValGly 388
      :::: :::: :::: ::::
388 euGInGlyHisArgGluProGlnProLeuHisAlaProProLeuAlaSer 404
      :::: :::: :::: ::::
321 CGAAGC.....GTGC 331
      |||||
405 ArgSerAspAsnArgGluGlyGlyProHisGlnArgAspTrpLeuG 421
332 CAACCTTAAGCGCGCAGCAAAATNNGCAATTCATCCATCCGCGTTGG 381
      :::: ::::
421 yArgAlaGlyGlnValArgGlyValArgGln.....ValArg.... 433
382 TGGAAGTCGGCTCGTANCCGTCGTCAGCAAAATCCCTGCCGTGCATGC 431
      ::|||
434 .....GlyValHisArgGluArgTrpArgGlyPro 443
432 CGAGCCGTTGCGCATCTTCGTCAATGCCATGGACACCAATCCCGTNGCG 481
      ||| |||||
444 ArgTyrValGlnHisLeuGlnArgAlaAspGlyArg..... 455
482 CAGACCCCTGGGTGTGATCAAGAAAGCCGCGANGATTTTCAGAGANGT 531
      |||||
456 .....CysGlyAla...ArgLeuProArgAlaLeuLeuArgLeuSerA 469
532 NTGCTGGTATTGAGCGG...TTTGACCGGACGGTAAATTCATGTCGTGTA 578
      ::||| |||||
469 IaGlyGlyTyrGluProArgGlyGlyHisAlaGlyAsnPro..... 482
579 GGCAGCTGGGCGAGAGCTGCCGCTGAAATGCTGCAACATCGAAGACAC 628
      |||||
483 .....GlnHisAsp..... 485
629 ATGAATTGCGGCGCCGCATCC..... 650
      ::|||
486 .....LysArgProArgTrpArgLeuGlnAspAspLysGluValArgG 500
651 .....GCCGCGTTTGAGTGG 665
      ::|||
500 IngLyLysGlyArgGlyPheProLeuArgGlyArgGlyArgAspAsnLeu 516
666 CAGCGACATTCATTTTCATGAGCGCGGTGCAAAACAAACCGTTTGG 715
      ::||| ::|||
517 rGlnGlnHisArg.....ArgCysLeuSerIleArgLeuG 528
716 CGATCAATTCACAGATGTAATGC.....CATC 744
      ::||| ::|||

```

[illegible]

COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
 MEDIUM TYPE: storage  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: IBM P.C. DOS 5.0  
 SOFTWARE: FastSeq for Windows 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/081.345  
 FILING DATE: Herewith  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/047,222  
 FILING DATE: May 20, 1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Warburg, Richard J.  
 REGISTRATION NUMBER: 32,327  
 REFERENCE/DOCKET NUMBER: 234/253  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (213) 489-1600  
 TELEFAX: (213) 955-0440  
 TELEX: 67-3510  
 INFORMATION FOR SEQ ID NO: 18:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 802 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-09-081-345-18

alignment scores:  
 Quality: 95.00 Length: 306  
 Ratio: 0.638 Gaps: 14  
 Percent Similarity: 48.693 Percent Identity: 21.895

alignment block:

US-09-303-518D-127/rev x US-09-081-345-18

Align seg 1/1 to: US-09-081-345-18 from: 1 to: 802

```

810 GACTTGAGAACCAACCAACCAACAT.....787
335 AspValylsThrThrAsnGlnHisSerLysGlnGlyAlaGluAlaGluSe 351
786 .....CAGCGCTGCTGCTTGCAGACGCGCTGTTGCAACCAACAGCT 747
351 rThGlyGlySerSerLeuGlyLeuAlaGlyThrSer.....Thm 364
746 CCGATGGCAATTCATCTTGATTAATGATGTCACAAAC.....GCT 706
364 etAsnAlaGlnGluGluLeuValLeuHisSerAlaLysSerSerProSer 380
705 TTGTGTTGACCGACCGCGCTCAATGAATGAATGCGTCC.....AC 662
381 PheAsnGlyLeuGlnLeuAsnGlyCysAsnAsnLysAlaValIleTh 357
661 TCAACCGCGCGCGATGCGCGCGCC.....6 636
397 rArgSnglGlnAlaArgAlaSerProValValGlyLProLeuGlnL 414
635 AATTCATGCTTTCGATGTTGGACGATTTTCAGACGCGACGCTGCGGC 586
414 yStyGlnSerLeuAspPheGlySerMetLeu.PheGly...SerCysPr 429
585 AGCTGCTTACACACATGATTTTACGCTGGTCAACAGGCTCAATACCA 536
429 oSerAlaLeuPro.....IleAsnThrAlaAspArgTyrHisAsnS 443
535 GC.....ANACNTGCTGTGAATTCNTGCGGCTCTTGTGATC... 498
443 eTlysGlyProValLysArgThrLysSerThrProPheGluLeuIleGln 459

```

```

497 ...ACAACCAAGGCTGTGCCGNAGCGA.....471
460 GlnAlaGlyThrAsnAspLeuAlaValGlyAspLysPheSerGlyLeuG 476
470 .....TTGGTGTCCATTCGATGACGAGAGA 446
476 uSerGlnLeuHisGlnHisTyrSerLeuAlaGlyLLeuGlnValGlnArgV 493
445 TGCGCAACGCGCTCGCATCGACGCGAGGATTTGCTGAACGAGCGAGNA 396
493 aAlaHisValSerSerGlnGluLeuAsnTyrSerLeuProGlyAlaCys 509
395 CGCAGCGCAGTC.....CACAAACGGATTTGATGATGATTCGANNON 355
510 AspAlaSerCysValProAlaGlnHisSerProGlyAlaLeuArgValHisLe 526
354 NANTTCNTGCGCGCTTAAGTTTGCACACGCTTCGGCGCGGTAGCTTGA 305
526 uTyrThrSerLeuAlaGlnAspProTyrPheSer.....537
304 ACTGATTTGCTGCTGCTTTCACGCGCATGACGACGACTGAAGTACG 255
538 .....SerSerProAsnSerAlaAspSerLys.....547
254 CGCTTTTCGCGCGGATGATGCGCGCGATTTGCTGTAACGCGCGGT 205
548 .....MetSerPheAspLeuProGlnLysGlnAspG 558
204 AACACCAACGCGCGATTCCTTTGCTTCAACAGCATTTGCGCTTTT 155
558 yAlaThrSerProGlyAlaLeuLeuProAlaSerSerThrThrSerPhe 575
154 TGACGCGATCGCTTCTTGTGACTTTCATCNAAGGGCGGCACACCGGCATAT 105
575 heTyrSerAsnProHisAspSerLeuValMetAsnThrLeuThrSerPhe 591
104 TCTTGCACAGACGCGCATTCGCTAATGACGCGCGCTCATTAATGAC 55
592 SerProProLeuAsnGlnGluThrAlaValGlnAlaProSerArgArgTh 608
54 TTGCTCGCGCTGCGC 39
608 rAspAspGlnIleLeuPro 613

```

seq\_name: /cgn2\_6/plodata/1/iaa/6A.COMB.pep:US-09-335-409-5  
 seq\_documentation\_block:  
 ; Sequence 5, Application US/09335409  
 ; Patent No. 6121029  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Schupp, Thomas  
 ; APPLICANT: Ligon, James  
 ; APPLICANT: Molnar, Istvan  
 ; APPLICANT: Zirkle, Ross  
 ; APPLICANT: Cyr, Devon  
 ; APPLICANT: Goerlach, Joern  
 ; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
 ; FILE REFERENCE: 4-30582A  
 ; CURRENT APPLICATION NUMBER: US/09/335,409  
 ; NUMBER OF SEQ ID NOS: 30  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 5  
 ; LENGTH: 7257  
 ; TYPE: PRT  
 ; ORGANISM: Sorangium cellulosum  
 ; US-09-335-409-5

alignment\_scores:  
 Quality: 93.50 Length: 538  
 Ratio: 0.408 Gaps: 29  
 Percent Similarity: 42.565 Percent Identity: 21.190

alignment\_block:  
US-09-303-518d-127 x US-09-335-409-5

Align seg 1/1 to: US-09-335-409-5 from: 1 to: 7257

```

61  TATGACGGGCGCGCATATCCGAAGTCGCGTGTGGCGAA...GATA 107
   :::::::::: :: :::::::::: :::::::::: ::
6356 pheasprgylglnalaprothrservalhstleuglyserleuaspgrl 6372
108  TCCCGGTATGCGCCCTTGATGAAAGTCAAGAGCGCATCCCTCAAA 157
   :::::::::: :: :::::::::: :::::::::: ::
6372 uargglyallleuaspralaaspralaprophaspralaaspraleuglug 6389
158  AAGGC.....CAAGTGTGTTGAGAACAAAGNAT 189
   :::::::::: :::::::::: :::::::::: ::
6389 luserleuvalarglycysasprervalleutprthralalaval 6405
190  CCGGGCGGTGTATACGCGGCN.....GTTTCAGCGCA 224
   :::::::::: :::::::::: :::::::::: ::
6406 AlaglyAlaglypheargasprproargleutprleuvalthrargl 6422
225  AATCGCGGCATCCATCGCGGGAAGCGCGTACTTCAGTCGCTC... 270
   :::::::::: :::::::::: :::::::::: ::
6422 yalaglnalalleglyAlaglyaspralservalalaglnalaproleul 6439
271  .....GTGATTGCCGTGAA..... 285
   :::::::::: :::::::::: :::::::::: ::
6439 euglyleuglyargvalillealaleugluhialaglyleuargcysala 6455
286  .....GGCAAC..GAGCAATGCA 302
   :::::::::: :::::::::: :::::::::: ::
6456 ArgilleasprleuaspProalaargargasprglylvalaspdlleue 6472
303  GTTCGAACGCTACGCGCGCGAAGCG.....TTGGCAACTTAAGCG 343
   :::::::::: :::::::::: :::::::::: ::
6472 ualaglyleuvalalaspaspralagluglvalalalphearglylg 6489
344  GCGANGAANTNGNNGCAATCTGATCCATCCGTTTGGACTGGCGTG 393
   :::::::::: :::::::::: :::::::::: ::
6489 lylgluargargvalalalargleuvalargargleuprogluthrasy 6505
394  CGTANCGTCCGCTTCAGCAAAATCCCGCGTATCGCGCGCTTCGC 443
   :::::::::: :::::::::: :::::::::: ::
6506 Arg.....GlulysillegluProalagluglyargProphear 6518
444  CATCTGTGTCAT.....GCGA 460
   :::::::::: :::::::::: :::::::::: ::
6518 glleugluileasprglyserglyValleuasprleuvalleuarglalat 6535
461  TGGACCAATCCGCTNCGCGCAGACCCCTGTGTGATCAAAAGAGCG 510
   :::::::::: :::::::::: :::::::::: ::
6535 hrgluarargargProproglyProglylvalglulilealavalgluala 6551
511  GNC.....GANGATTTCAGACGANGTCTGCTGATTCGAGCGC 548
   :::::::::: :::::::::: :::::::::: ::
6552 AlaglyleuasprpheleuasprvalmetarglaketglyletryProgl 6568
549  TTTGACGAGCGGTAAATTCATGTG.....TGTAG..... 579
   :::::::::: :::::::::: :::::::::: ::
6568 yProglyasprglyProvalalaleuglyalaglycysserglyArgllev 6585
580  ..GCAGCTGGCGCAGACGTCGCT.....GAAATGTCGCC 615
   :::::::::: :::::::::: :::::::::: ::
6585 alAlametylgluglyvalglusertleuarglileglylinspralval 6601
616  AACATCGAAACATGAAATTCGCG..... 639
   :::::::::: :::::::::: :::::::::: ::
6602 AlavalAlalaproheserphelglythrhlsvalthrileaspralar 6618
640  ....GGCCGCGATCCGCGGTTTGAAGTGGCAGCGACATTCATTCAATG 685
   :::::::::: :::::::::: :::::::::: ::
6618 tleualaproargProalalalaleutThrAlalaglnalalealaleu. 6634

```

```

686  AGCCGGTCGTGCAACAAACCGTTTGACCATTCATTCAGATGTA 735
   :::::::::: :::::::::: :::::::::: ::
6635 ..Proval..AlapheethrAlatrp.....Tyrglyleu 6645
736  ATTGCCATCGGACGCTTTGTTGCACAGCGCGTCTGACACGACGCGCT 785
   :::::::::: :::::::::: :::::::::: ::
6646 valhstleuglyargleuargla.....glylulargya 6657
786  GATT.....GCTTGGGTGTTCTCAAGTCAACAAACCGCGCT 826
   :::::::::: :::::::::: :::::::::: ::
6657 lleuilehsserlatrhrglygluthrglyleualalavalglnle 6674
827  TCCGTCACGCTTTGGTGCAGAGA..... 852
   :::::::::: :::::::::: :::::::::: ::
6674 laarghls...leuglyalagluileprialatrhralaglythrProglu 6689
852  ..... 852
6690 lysArglatrpleuargluglinglelyllealanhlsvalmetasprer 6706
853  .....TGGCAAT.....ACTGCGGCGCAATTGG 877
   :::::::::: :::::::::: :::::::::: ::
6706 gserleuasprphealagluglnvalleualalathrlsyglugly 6723
878  TTGACGACAGACAACCGCGTATTCGCGTTCGATTCGACGCGCGCAT 927
   :::::::::: :::::::::: :::::::::: ::
6723 alaspvalalleuasnserserleuarglyAlalalaleaspralaser 6739
928  ACACAGAGCGCGCAGATTAATTGGAGCGCTCCACATTCAGATTCCGT 977
   :::::::::: :::::::::: :::::::::: ::
6740 Ser.....Thle 6742
978  TATCGAAGAGCGCGCAGCAAGAGCTGTTCGCTGGTTCGCGCGCAGC 1027
   :::::::::: :::::::::: :::::::::: ::
6742 uvalproasprglyArgphellegluu.....glyl 6754
1028  CGGACAAATCTCCATCAGCGGTACGACCCCTGGCATTCCTGAANAAC 1077
   :::::::::: :::::::::: :::::::::: ::
6754 hrasprietyrAlalaspargserleuglyleualhstphearglyser 6770
1078  AAACCTTCAGATTCAACGACGCGGTCAACGCGCGCGCGCATGTGT 1127
   :::::::::: :::::::::: :::::::::: ::
6771 leusertyr.....SerAlaval.....AspleuAlaglyle 6781
1128  GCCGATTGTACTTACGAGCGCGTA..... 1152
   :::::::::: :::::::::: :::::::::: ::
6781 ualavalargargProgluargvalAlalalaleuAlagluvala 6798
1153  .....ATGCCGTAGACATCCGTGCT 1173
   :::::::::: :::::::::: :::::::::: ::
6798 spleuAlalarglyAlaleuglnProleuProvalgluilepPro 6814
1174  ACCCTGCTTTGGCGGATTAATCGTCGCGATACCGACGCGCGAC 1223
   :::::::::: :::::::::: :::::::::: ::
6815 leuserarglAlalasprialphe..ArglymetAlaglnalaglnhls 6830
1224  ATTGGGTGCTTGGAAATTCGACGAAAGACCTGCTTTCGACGCTTGC 1273
   :::::::::: :::::::::: :::::::::: ::
6831 leuglylsleuvalleualaleugluaspProasp...Valarglilear 6846
1274  TCTGCGCGCGGCA 1286
   :::::::::: :::::::::: :::::::::: ::
6846 gvalProglylu 6850

```

seq\_name: /cgn2\_6/prodata/1/laa/5B\_COMB.pep:US-09-568-102-5

seq\_documentation\_block:

```

; Sequence 5, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schnupp, Thomas
; APPLICANT: Ligon, James

```

```

APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOCHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/568,102
CURRENT FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: 09/335,409
PRIOR FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: patentln Ver. 2.0
SEQ ID NO 5
LENGTH: 7257
TYPE: prt
ORGANISM: Sorangium cellulosum
us-09-568-102-5

```

## alignment\_scores:

```

Quality: 93.50      Length: 538
Ratio: 0.408      Gaps: 29
Percent Similarity: 42.565      Percent Identity: 21.190

```

## alignment\_block:

```

US-09-303-518D-127 x US-09-568-102-5 ..

```

```

Align seg 1/1 to: US-09-568-102-5 from: 1 to: 7257

```

```

61  TATGACGGCCGCTGATACGAGATGCGTGTGGCGAA...GAATA 107
   :::::  :::::  :::::  :::::  :::::  :::::
6356 PheasprglnAlaProthrsValHisLeuGlySerleuasp1 6372
108  TGGCGGATGCGCCCTNGATGAAAGTCAAGAGCGATGCGTCAAA 157
   :::::  :::::  :::::  :::::  :::::  :::::
6372 uatgilyvalleuaspAlaSprAlaProthrsAlaSprAlaLeuGly 6389
158  AAGC.....CAAGTGTGTGTAAGACAAAGAAAGNT 189
   :::::  :::::  :::::  :::::  :::::  :::::
6389 luseileuValArgGlyCysAspSerValLeuThrValGlnAlaVal 6405
190  CCGGCGTGTGTACCGCCN.....GTTTCAGCGAA 224
   :::::  :::::  :::::  :::::  :::::  :::::
6406 AladilyAlaGlyPheArgAspProArgLeuThrValThrArg1 6422
225  AATGCGCGCATCCATCGCGGAAAGCGCGTACTGAGTGGTC... 270
   :::::  :::::  :::::  :::::  :::::  :::::
6422 yAlaGlnAlaIleGlyAlaGlyAspValSerValAlaGlnAlaProLeu 6439
271  .....GTGATTCGCGTGAA..... 285
   :::::  :::::  :::::  :::::  :::::  :::::
6439 euGlyLeuGlyArgValIleAlaLeuGlnHisAlaGlyLeuArgCysAla 6455
286  .....GGCAAC...GACGAATGCA 302
   :::::  :::::  :::::  :::::  :::::  :::::
6456 ArgIleAspLeuAspProAlaArgArgAspGlyValAspGlyLeuLe 6472
303  GTTCGAACGCTACGCGCCGAGCG.....TTGGCAAACTTAAGCG 343
   :::::  :::::  :::::  :::::  :::::  :::::
6472 uAlaGlyLeuAlaAspAspAlaGlnGlyValAlaIleArgGlyG 6489
344  GCGAAGAAATNNNGCAATCTGATCAATCGGTTGTGAGTCGCGCTG 393
   :::::  :::::  :::::  :::::  :::::  :::::
6489 lYgluArgArgValAlaArgLeuValArgArgLeuProGluThrAspCys 6505
394  CSTANCCGTCGTCAGCAAAATCCCTCGCGCATGCGCGCGCTGCC 443
   :::::  :::::  :::::  :::::  :::::  :::::
6506 Arg.....GluysIleGlnProAlaGlnGlyArgProPheAr 6518
444  CATCTTCGCAAT.....GCGA 460
   :::::  :::::  :::::  :::::  :::::  :::::
6518 gLeuGlnIleAspGlySerGlyValLeuAspAspLeuValLeuArgAla 6535

```

```

461  TGGACCAACATCCGCTNCGCGACACCCCTGGTGTGATCAAGAACCC 510
   :::::  :::::  :::::  :::::  :::::  :::::
6535 hrGlnArgArgProProGlyValGlnIleAlaGlnAla 6551
511  GNC.....GANGATTTCAGACGANGTGTGTGATTCAGCCG 548
   :::::  :::::  :::::  :::::  :::::  :::::
6552 AlaglyLeuAsnPheLeuAspValMetArgAlaMetGlyIleThrProG 6568
549  TTGACCGAGCGTAATCCATGTG.....TGTAG..... 579
   :::::  :::::  :::::  :::::  :::::  :::::
6568 yProGlyAspGlyProValAlaLeuGlyAlaGlyCysSerGlyArgTle 6585
580  ..GCACTGCGCGACAGTGCCT.....GAAATGCTGCC 615
   :::::  :::::  :::::  :::::  :::::  :::::
6585 AlAlaMetGlyGlyValGlnSerLeuArgIleGlyGlnAspValAl 6601
616  AACATCGAAACATGAAATCGGC..... 639
   :::::  :::::  :::::  :::::  :::::  :::::
6602 AlAlaValAlaProPheSerPheGlyThrHisValThrIleAspAlaArg 6618
640  ....GCCCCGATCCGCGCGTTTGAAGTGCAGCAGCATTCATTCATTCG 685
   :::::  :::::  :::::  :::::  :::::  :::::
6618 tLeuAlaProArgProAlaIleuThrAlaIleGlnAlaIleAla 6634
686  AGCGGTCGCTGCAACAAACCGTTTGGACCATTCATTCAGAGATGA 735
   :::::  :::::  :::::  :::::  :::::  :::::
6635 ..ProVal...AlaPheMetThrAlaTrp.....TyrGlyLeu 6645
736  ATTGCGATCCGAGCTTGTTCGACAGGCGCGTGAACACGAGCGCGT 785
   :::::  :::::  :::::  :::::  :::::  :::::
6646 ValHisLeuGlyArgLeuArgAla.....GlyGlnArg 6657
786  GATT.....GTTTGGTGTGTTCAGTCAACAAACCGCGCT 826
   :::::  :::::  :::::  :::::  :::::  :::::
6657 lLeuIleHisSerAlaThrGlyGlyThrGlyLeuAlaValGlnIleA 6674
827  TGGGTACCGCTTGTGGTGGCAAGTA..... 852
   :::::  :::::  :::::  :::::  :::::  :::::
6674 lArgGlns...LeuGlyAlaGlnIlePheAlaThrAlaGlyThrProGlu 6689
852  ..... 852
6690 LysArgAlaThrLeuArgGlnGlnGlyIleAlaHisValMetAspSerArg 6706
853  .....TGCAGATT.....ACTCGCGGCAATTCG 877
   :::::  :::::  :::::  :::::  :::::  :::::
6706 gSerLeuAspPheAlaGlnGlnValLeuAlaIleThrIysGlyGlnGly 6723
878  TTGACGACAGACAAACCGGTGATTCGCGTGGATTCGAGTTCAGCGCGAT 927
   :::::  :::::  :::::  :::::  :::::  :::::
6723 AlAspValValLeuAsnSerLeuSerGlyAlaIleAlaIleAspAlaSerLeu 6739
928  ACACAAAGCGCGCACGATTATTGGGACGCTACCAATCAATGATTTCGT 977
   :::::  :::::  :::::  :::::  :::::  :::::
6740 Ser.....Thle 6742
978  TATGAGAAGAGCGCGCAAGAGAGCTGTGCGTGGTGGCGCGGAGC 1027
   :::::  :::::  :::::  :::::  :::::  :::::
6742 uValProAspGlyArgPheIleGlnLeu.....GlyLysT 6754
1028  CGGCAAAATACCTCCATCCAGCGGTACAGACCTCGGCAATTCCTGAAAAC 1077
   :::::  :::::  :::::  :::::  :::::  :::::
6754 hrAspIleThrAlaAspArgSerLeuGlyLeuAlaHisPheArgLysSer 6770
1078  AAACCTTCAGAGTTCAGAGACCGCTCAAGAGGTGGCGCGCGCGCAATGG 1127
   :::::  :::::  :::::  :::::  :::::  :::::
6771 LeuSerTyr.....SerAlaVal.....AspLeuAlaGlyLe 6781
1128  GCCGATTGGTACTTACAGCGCGTA..... 1152
   :::::  :::::  :::::  :::::  :::::  :::::
6781 uAlaValArgArgProGluArgValAlaAlaLeuLeuAlaGlnValAla 6798
1153  .....ATGCGCGTACAGATCTGCCT 1173

```



```

6723 aalspvalleuasnserleuserglyalalaalleasplasleu 6739
928 ACACAAGGCGCCGCGATTATTTGGACGCTACACAAATCAGATTTCGT 977
6740 Ser.....Thre 6742
978 TATCGAAGAGCGCCGACGAAAGCTGTTCGGCTGGTGGCGCCGAC 1027
6742 uvalproaspglyargpheileu.....gly 6754
1028 CGGACAATATCTCATCGACGCGTACGACCTCGCCATTTCGAAAGC 1077
6754 hrspileyrallaspargserleuglyleualhisphearglyser 6770
1078 AACTCTTCAGTTCCAGACAGCGCTCAACGGTGGCGACCGCCCATGT 1127
6771 leuser Tyr.....seralaval.....aspleualiglyle 6781
1128 GCGGATTGGTACTTACGACGCGTA..... 1152
6781 ualavalargargprogluargvalalalaaleuualagluvala 6798
1153 .....ATGCGCTAGACATCGTCCT 1173
6798 spleuualargglyalaleuulinproleuprovaleuilephepro 6814
1174 ACCCTCTTTGGCGGATTAAATCGTGGCGATACCGACGCGCCAGC 1223
6815 leuserargalalalaspalphe..Arglymetalaglinalaglnhis 6830
1224 ATTGGTTGCTTGGAATTGGACGAGAAAGACTCGCTTGTGCACCTCG 1273
6831 leuglylysleuvalleualleuualinsuproasp...Valargleuar 6846
1274 TCTGCCCGGGGCAA 1286
6846 gyalproglygu 6850

```

seq\_name: /cgm2\_6/ptodata/1/iaa/6B\_COMB pep: US-09-568-480-5

```

seq_documentation_block:
; Sequence 5, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zitzke, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-480-5

```

```

alignment_scores:
Quality: 93.50 Length: 538
Ratio: 0.408 Gaps: 29
Percent Similarity: 42.565 Percent Identity: 21.190

```

alignment\_block:  
US-09-303-518D-127 x US-09-568-480-5 ..

Align seg 1/1 to: US-09-568-480-5 from: 1 to: 7257

```

61 TATGACGGGCGCCGATTCAGAGTCGCGTGTGGCGAA...GATA 107
6356 Pheaspglylinalaaprothservalhisleuylserleuasp 6372
108 TGCCGATATGCGCCCTNGATGAAGTCAAGAGGCGATGCCGTA 157
6372 uargglyvalleuaspalaspalapropheaspalasleuclug 6389
158 AAGC.....CAAGTCTGTTGAGCAAAAAGNAT 189
6389 Insleuvalarglycysaspservalleurtrhrvalalaval 6405
190 CCGGCGGCGGTGTACCGGCCN.....GTTCCAGGCAA 224
6406 Alaglyalaglypheargasproproargleutripheualthrarg 6422
225 AATGCGCGCCATCCATCGCGCGGAAAGCGGACTTCACTCGCTC... 270
6422 yalaglinalleuglyalaglyaspvalservalalaglinalproleu 6439
271 .....GTGATTCGCGTTGA... 285
6439 euuglyleuglyargvalillealaleuulinisalaglualeuargysala 6455
286 .....GGCAAC...GAGCAAAATCGA 302
6456 Argllaspheuasproalalargargspglyglivalaspduleue 6472
303 GTTCGAAGCTTACGCGCGCCAGACG.....TTGGCAAACTTAAGC 343
6472 ualagluleuualaspaspalagluaglivalalapheargly 6489
344 GCGANSAANTNNGNCAATCTGATCCATCCGCTTGTGCATGGCGTC 393
6489 lycluarargvalalalargleuvalargleuproglythrasy 6505
394 CGTANCCGTCGTTCCAGCAAAATCCCTGCGGATGCGGCGGCTTCC 443
6506 Arg.....glutylleugluproalagluylargprohear 6518
444 CATCTTCGTCAAT.....GCCA 460
6518 gleuglilleaspglyserglyalleuaspaspleuallleuargalat 6535
461 TGGACACCAATCCGCTNCGCGGACAGACCTGCTGTGATCAAGAGCC 510
6535 hrgluarargproproglyproglygluvalisluilealavalgluala 6551
511 GNC.....GANGATTTACAGACGANGTGTGATTTAGCGG 548
6552 Alaglyleuasnpheluaspvalmetargalameglylleytrprogl 6568
549 TTTGACCGAGCGTAAATCCATGTG.....TGTAAG..... 579
6568 yproglyaspglyprovalalaleuuglyalaglyssserglyarglleav 6585
580 ..GCAGTGGCGGACGACGTCGCTC.....GAAATGCTGCC 615
6585 alalameuglygluylgluileuargllleuglylinspvala 6601
616 AACATCGAAACACATGATTCGCG..... 639
6602 Alavalalaprotheserpheeglythrhisvalthrlleaspalargme 6618
640 ....GGCCGCGATCCGCGGTTTGAAGTGCACGCAATTCATTTCATG 685
6618 lleualaproargproalalaleuthralalaglnalalaleu. 6634
686 AGCCGCTCGTGCAAAACACCGTTTGGACCATCAATATCAAGATGA 735
6635 ..Proval...AlaphemetthrAlatrp.....Tyrglyleu 6645

```



```

736 ATTGCATCGACGTTTGTTCGAACAGCCGCTTGAACACCGACGCGT 785
    ::::::::::::::::::::
6646 VAIHSLLEUGLYARGLEUARGLA.....GLYLUARGVA 6657
786 GATT.....GCTTGGGGGTTCACAGTCAACAACACCGCCCT 826
    ::::::::::::::::::::
6657 LLEUHLHSLSERIALATHRGLYGlyThGlyLeuAlaValaGlnIleA 6674
    ::::::::::::::::::::
827 TCGGTACCGTTTGGGTGCGAAGTA..... 852
    ::::::::::::::::::::
6674 lAaRgHs..LeuGlyAlaGlnIlePheAlaThrAlaGlyThrProGlu 6689
    ::::::::::::::::::::
852 ..... 852
6690 LysArgAlaThrPLeuArgGlnGlnIleAlaHisValMetAspSerAr 6706
853 .....TCGCAAT.....ACTGGGGCGAATTGG 877
6706 gSerLeuAspPheAlaGlnGlnValLeuAlaAlaThrLysGlyGlnGly 6723
    ::::::::::::::::::::
878 TTGACGACAGACACCGCGTATTTCGGTTCGTAATTGAACGGCGCAT 927
    ::::::::::::::::::::
6723 alAspAlValLeuAsnSerLeuSerGlyAlaAlaIleAspAlaSerLeu 6739
928 ACACAAAGCGCGCAGATTATTGGACCGCTACACAAATCAGATTCCGT 977
    ::::::::::::::::::::
6740 Ser.....Thre 6742
978 TATGACAAAGACCGCGACAAAGCGTTCGGTGGTTCGGCGCGCAGC 1027
    ::::::::::::::::::::
6742 uAlaProAspGlyArgPheIleGluLeu.....GlyLysT 6754
1028 CGGACAAATATCTCCATCAGCGGTACGACCCGCGCATTTCTGAAAAC 1077
    ::::::::::::::::::::
6754 hrAspLierYrAlaAspArgSerLeuGlyLeuAlaHisPheArgLysSer 6770
1078 AAATCTTCAGTTCACGACAGCGCGTCAACGGTGGACCGCGCATTTGT 1127
    ::::::::::::::::::::
6771 LeuSerYr.....SerAlaVal.....AspLeuAlaGlyLe 6781
1128 GCGCATTTGTAATTACGAGCGCGTA..... 1152
    ::::::::::::::::::::
6781 uAlaValArgArgProGlnArgValAlaAlaLeuLeuAlaGlnValAla 6798
1153 .....ATGCCGCTAGACATCTGCCT 1173
6798 splLeuAlaArgGlyAlaLeuGlnProLeuProValGlnIlePhePro 6814
1174 ACCCTGCTTTGGCGATTAAATCGTGGGATACGACGACGCGCAGC 1223
    ::::::::::::::::::::
6815 LeuSerArgAlaAlaAspAlaPhe..ArgLysMetAlaGlnAlaGlnHis 6830
1224 ATTGGTGTGCTTGGAATTGACGAAGAAGACCTCGCTTGTGACGCTTG 1273
    ::::::::::::::::::::
6831 LeuGlyLysLeuValLeuAlaLeuGlnAspProAsp...ValArgIleAr 6846
1274 TCTGCCCGGGCAA 1286
    ::::::::::::::::::::
6846 gValProGlyGln 6850
seq_name: /cgn2_6/ptodata/1/laa/GB_COMB.pep:us-09-568-486-5
seq_documentation_block:
; Sequence 5, Application US/09568486
; Patent No. 635459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern

```

```

; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentl Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
; US-09-568-486-5

```

```

alignment_scores:
  Quality: 93.50      Length: 538
  Ratio: 0.408      Gaps: 29
  Percent Similarity: 42.565      Percent Identity: 21.190

```

```

alignment_block:
US-09-303-518d-127 x US-09-568-486-5

```

```

Align seq 1/1 to: US-09-568-486-5 from: 1 to: 7257

```

```

61 TATGACGGCGCCGTCATTACCGAAGTCGCTTGGCGAA...GATA 107
    ::::::::::::::::::::
6356 PheAspGlyGlnAlaProThrSerValValHisLeuGlySerLeuAspG 6372
108 TGCCGATATGCGCCCTNGATGAAAGTCAGAAAGCGGATGCGTCAAAA 157
    ::::::::::::::::::::
6372 uArgGlyValLeuAspAlaAspAlaProPheAspAlaLeuGln 6389
158 AAGGC.....CAAGTGTGTGAAGACAAAGNAT 189
    ::::::::::::::::::::
6389 LuSerLeuValArgGlyCysAspSerValLeuTrpThrValGlnAlaVal 6405
190 CGGGCGTGTGTTTACCGCCCN.....GTTTCAGCGCAA 224
    ::::::::::::::::::::
6406 AlaGlyAlaGlyPheArgAspProArgLeuTrpLeuValThrArgG 6422
225 AATCGCCCGCATCCATCGCGCGCAAGCGCGTACTTACGCGGTC.... 270
    ::::::::::::::::::::
6422 yAlaGlnAlaIleGlyAlaGlyAspValSerValAlaGlnAlaProLeu 6439
271 .....GTGATTGGCGCTTGAA..... 285
6439 euGlyLeuGlyArgValIleAlaLeuGlnHisAlaGlnLeuArgCysAla 6455
286 .....GGCAAC...GACGAATGCA 302
6456 ArgIleAspLeuAspProAlaArgArgAspGlyGlnValAspGlnLeu 6472
303 GTTCGACACCTTACGGCGCGAAGCG.....TTGCAAACTTAACG 343
    ::::::::::::::::::::
6472 uAlaGlnLeuLeuAlaAspAspAlaGlnGlnValAlaAlaPheArgGly 6489
344 GCGANGAANTNNGCATTCATTCATCCGATTGTGACGTCGCGTG 393
    ::::::::::::::::::::
6489 LysGlnArgArgValAlaArgLeuValArgArgLeuProGlnThrAspCys 6505
394 CGTANCGTCCGTTCAAGCAAAATCCCTGCCCTCGATGCCGACGCGTGGC 443
    ::::::::::::::::::::
6506 Arg.....GlnLysIleGluProAlaGlnGlyArgProPheAr 6518
444 CATCTGTCGTCAT.....GGCA 460
    ::::::::::::::::::::
6518 gLeuGlnIleAspGlySerGlyValLeuAspAspLeuValLeuArgAlaAr 6535
461 TGACACCAATCGCTNCGCGCAGACCGCTGTTGTGATCAAGAAGCC 510
    ::::::::::::::::::::
6535 hrGlnuArgArgProProGlyProGlyGlnValGlnIleAlaValGlnAla 6551

```

```

511 GNC.....GANGATTTCAGACGANGTNTGCTGATTGAGCCG 548
    |||
5552 AlaGlyLeuAspPheLeuAspValMetArgAlaMetGlyIleTyrProG1 6568
    |||
549 TTGACCGAGCGGATAATCCATGTG.....TGTAAG..... 579
    |||
6568 yProG1AspGlyProValAlaLeuG1ValaG1LucSerSerGlyArgIleV 6585
    |||
580 ..GCAGCTGGCGCAGACGTCGCGTCT.....GAAATGCTGCC 615
    |||
6585 aAlaMetG1yG1uG1yValaG1uSerLeuArgIleG1yG1uAspValVal 6601
    |||
616 AACATCGAAGACATGATTCGCG..... 639
    |||
6602 AlaValAlaProPheSerPheG1yThrHisValThrIleAspAlaArgMe 6618
    |||
640 ...GGCCCGCATCCGCGCGGTTTGAGTGGCAGCAGCATTCATTTCATG 685
    |||
6618 tLeuAlaProAlaProAlaAlaLeuThrAlaAlaG1nAlaAlaLeu. 6634
    |||
686 AGCCGGTGGTGGCAAAACCGCTTGGACCATCATCATGATGATGTA 735
    |||
6635 ..ProVal...AlaPheMetThrAlaTrp.....TyrGlyLeu 6645
    |||
736 ATTGGCATCGGACGCTTTGTCGACAGCGCGCTGTAACACCGCGCGCT 785
    |||
6646 ValHisLeuG1yArgLeuAlaArgAla.....GlyG1uArgVa 6657
    |||
786 GATT.....GCTTGGGGTGTCTCANGTCACAAACACGCGCT 826
    |||
6657 lLeuIleHisSerAlaThrG1yG1yThrG1yLeuAlaAlaValaG1nIleA 6674
    |||
827 .TGCGTACCGTTTGGTGGCAAGTA..... 852
    |||
6674 lArgHis...LeuG1yAlaG1uIlePheAlaThrAlaG1yThrProG1u 6689
    |||
852 ..... 852
    |||
6690 LysArgAlaTrpLeuArgG1uG1uG1yAlaHisValMetAspSerAr 6706
    |||
853 .....TCGCATAT.....ACTGGCGGCAATGG 877
    |||
6706 gSerLeuAspPheAlaG1uG1uValaLeuAlaAlaThrLysG1yG1uG1yV 6723
    |||
878 TTGACGAGACACACCGCGGTATTCGCTGATTGAACGCGCGGATT 927
    |||
6723 aAspValValLeuAsnSerLeuSerG1yAlaAlaIleAspAlaSerLeu 6739
    |||
928 ACACAGCGCGCGCATTTTGGGACGCTACACACATGAGATTTCGCT 977
    |||
6740 Ser..... 977
    |||
978 TATCGAAGAAGGCGCAGCAAGAGCTGTGCGTGGTGGCGCGCAGC 1027
    |||
6742 uValProAspG1yArgPheIleG1uLeu.....GlyLysT 6754
    |||
1028 CGNACAATATCTCCATCAGCGCGTACACACCTCGCCATTTCCTGAAAC 1077
    |||
6754 hrAspIleTyrAlaAspArgSerLeuG1yLeuAlaHisPheArgIlySer 6770
    |||
1078 AAACCTTTCAAATTCACGACGCGCTCAACGCGTGGCGCGCGCATGTG 1127
    |||
6771 LeuSerTyr.....SerAlaVal.....AspLeuAlaG1yLe 6781
    |||
1128 GCGGANTGGTACTTACGAGCGCGTA..... 1152
    |||
6781 uAlaValAlaArgArgProG1uArgValAlaAlaLeuLeuAlaG1uVala 6798
    |||
1153 .....ATGCCGTAGACATCTCTGCGCT 1173
    |||
6798 sPleuLeuAlaArgG1yAlaLeuG1nProLeuProValG1uIlePhePro 6814
    |||
1174 ACCCTGCTTTTCCGCGATTATATCGTGGCGATACGACAGCGCGCAC 1223

```

```

seq_name: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:US-09-568-472-5

```

```

seq_documentation_block:

```

```

; Sequence 5, Application US/09568472

```

```

; Patent No. 6358719

```

```

; GENERAL INFORMATION:

```

```

; APPLICANT: Schupp, Thomas

```

```

; APPLICANT: Ligon, James

```

```

; APPLICANT: Molnar, Istvan

```

```

; APPLICANT: Zirkle, Ross

```

```

; APPLICANT: Cyrt, Devan

```

```

; APPLICANT: Goerlach, Joern

```

```

; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

```

```

; FILE REFERENCE: 4-30582A

```

```

; CURRENT FILING DATE: 2000-05-10

```

```

; PRIOR APPLICATION NUMBER: 09/335,409

```

```

; NUMBER OF SEQ ID NOS: 30

```

```

; SOFTWARE: PatentIn Ver. 2.0

```

```

; SEQ ID NO 5

```

```

; LENGTH: 7257

```

```

; TYPE: PRT

```

```

; ORGANISM: Sorangium cellulosum

```

```

US-09-568-472-5

```

```

alignment_scores:

```

```

Quality: 93.50 Length: 538

```

```

Ratio: 0.408 Gaps: 29

```

```

Percent Similarity: 42.565 Percent Identity: 21.190

```

```

alignment_block:

```

```

US-09-303-518d-127 x US-09-568-472-5 ..

```

```

Align seg 1/1 to: US-09-568-472-5 from: 1 to: 7257

```

```

61 TATGAGGCGCGCGCATTAACGAGTGGCTTGGTGGCGAA...GATA 107

```

```

6356 PheAspGlyGlnAlaProThrSerValValHisLeuGlySerLeuAspG1 6372

```

```

108 TGCCGGTATGCGCCCTGATGAAAGTCACGAAGAGCGGATGCCGCAAA 157

```

```

6372 uArgG1yValLeuAspAlaAspAlaProPheAspAlaAspAlaLeuG1u 6389

```

```

158 AAGC.....CAAGTGTGTTGAAGACAAAGAT 189

```

```

6389 lSerLeuValAlaArgG1yCysAspSerValLeuTrpThrValGlnAlaVal 6405

```

```

190 CCGGCGGTGGTGTTCACCGCGCN.....GTTTCAGGCAA 224

```

```

6406 Alag1yAlaG1yPheArgAspProArgLeuTrpLeuValThrArgG1 6422

```

```

225 AATGCCCGCATTCATCGCGCGGCAAGCGGTACTTCAGTCGCTC... 270

```

```

6422 yAlaGlnAlaIleG1yAlaG1yAspValSerValAlaGlnAlaProLeu 6439

```

```

271 .....GTGATGCGCTTGA... 285

```

```

6439 euG1yLeuG1yArgValIleAlaLeuG1uHisAlaG1uLeuArgCysAla 6455

```

```

286 .....GGCAAC...GACGAATGCA 302

```

```

6456 ArgIleAspLeuAspProAlaArgArgSpLyIuValAspIuLeuLe 6472
      |||:|:| |||:|:|:|
303 GTTCGAACGCTAGCGCCCGAAGCG.....TTGGCAACTTAAGCG 343
      ||| ||| |||:|:|
6472 uAlaGluLeuLeuAlaAspAlaGluGluAlaAlaPheArgGlyG 6489
      ||| ||| |||:|:|
344 GCGANGAANNNGNNGCAATCTGATCCATCCGTTGTGACGCGCGCTG 393
      ||| ||| |||:|:|
6489 LyGluArgArgValAlaArgLeuValArgArgLeuProGluThrAspCys 6505
      ||| ||| |||:|:|
394 CGTACCGCGCTGTCACCAAAATCCCTGCGGTGATGCGCGCGCTGCG 443
      ||| ||| |||:|:|
6506 Arg.....GluLysIleGluProAlaGluGluArgProPheArg 6518
      ||| ||| |||:|:|
444 CATCTGTCTCAAT.....GCGA 460
      ||| ||| |||:|:|
6518 GluGluIleLeuAspLySerGlyValLeuAspAspLeuValLeuArgAla 6535
      ||| ||| |||:|:|
461 TGGACACCAATCGCTGCGCGCGAGACCGTGTGATCAAGAGAGCG 510
      ||| ||| |||:|:|
6535 hrGluArgArgProProGlyProGlyLysIleAlaValGluAla 6551
      ||| ||| |||:|:|
511 GNC.....GANGATTCAGACGANGTNTGCTGTGATTTGACCG 548
      ||| ||| |||:|:|
6552 AlaGlyLeuAsnPhenLeuAspValMetArgAlaMetGlyIleTyProG 6568
      ||| ||| |||:|:|
549 TTTGACCGACGCTAAATCCATGTG.....TGTAAG..... 579
      ||| ||| |||:|:|
6568 yProGlyAspLyProValAlaLeuGlyAlaGluCysSerGlyArgIleV 6585
      ||| ||| |||:|:|
580 ..GCAGTGGCGCAGACGTGCCGCT.....GAAATGTCGCC 615
      ||| ||| |||:|:|
6585 alAlaMetGlyGluGlyAlaGluSerLeuArgIleGlyGlnAspVal 6601
      ||| ||| |||:|:|
616 AACATCGAACAACATGAATTCGCG..... 639
      ||| ||| |||:|:|
6602 AlaValAlaProPheSerPheGlyThrHisValThrIleAspAlaArg 6618
      ||| ||| |||:|:|
640 ....GCGCGCGATCGCGCGGTTTGAGTGCAGCGACACATTATTCAT 685
      ||| ||| |||:|:|
6618 tLeuAlaProArgProAlaAlaLeuThrAlaAlaGlnAlaAlaLeu 6634
      ||| ||| |||:|:|
686 AGCGCGTGGTGCACAAACCGTTGACCATCATTAACAAGTGA 735
      ||| ||| |||:|:|
6635 ..ProVal...AlaPheMetThrAlaTrp.....TyGlyLeu 6645
      ||| ||| |||:|:|
736 ATTGCCATCGGACGTTGTTGCAACAGCGCGTGAACACCGAGCGCT 785
      ||| ||| |||:|:|
6646 ValHisLeuGlyArgLeuArgAla.....GlyGluArgVal 6657
      ||| ||| |||:|:|
786 GATT.....GCTTGGGTGGGTTCACAGTCAACAACACCGCCCT 826
      ||| ||| |||:|:|
6657 tLeuIleHisSerAlaThrGlyGlyThrGlyLeuAlaAlaValGlnIle 6674
      ||| ||| |||:|:|
827 TGCCTACCGCTTTGGGTGCGAAGTA..... 852
      ||| ||| |||:|:|
6674 laArgHis...LeuGlyAlaGluIlePheAlaThrAlaGlyThrPro 6689
      ||| ||| |||:|:|
852 ..... 852
      ||| ||| |||:|:|
6690 LysArgAlaTrpLeuArgGluGlnGlyLeaHisValMetAspSer 6706
      ||| ||| |||:|:|
853 .....TCGCAAT.....ACTCGGCGCAATGG 877
      ||| ||| |||:|:|
6706 gSerLeuAspPheAlaGluGlnValLeuAlaAlaThrLysGlyGluG 6723
      ||| ||| |||:|:|
878 TTGACGACAGACAACCGCGTATTCGCTGCGTATTAAGCGCGCGATT 927
      ||| ||| |||:|:|
6723 alAspValValLeuAsnSerLeuSerGlyAlaAlaIleAspAlaSer 6739
      ||| ||| |||:|:|
928 ACACAAGGCGCGACGATTATTGGAGCGCTACCAACAATGATTCCGT 977
      ||| ||| |||:|:|

```

```

6740 Ser.....ThrLe 6742
      ||| ||| |||:|:|
978 TATCGAAGAGCGCCGACGAAGAGCTGTGCGCTGGGTGGCGCGACG 1027
      ||| ||| |||:|:|
6742 uValProAspGlyArgPheIleGluLeu.....GlyLysT 6754
      ||| ||| |||:|:|
1028 CGACAAATATCTCATCACGCGCTAGACCGTGGCCATTTCCTGAAAAC 1077
      ||| ||| |||:|:|
6754 hrAspIleTyAlaAspArgSerLeuGlyLeuAlaHisPheArgLysSer 6770
      ||| ||| |||:|:|
1078 AAATCTTCAGTTTCAGACAGCGCTCAAGCGTGGCGACCGCGCATGT 1127
      ||| ||| |||:|:|
6771 LeuSerTy.....SerAlaVal.....AspLeuAlaGlyLe 6781
      ||| ||| |||:|:|
1128 GCGATGTGTACTACGACGCGTA..... 1152
      ||| ||| |||:|:|
6781 uAlaValArgArgProGluArgValAlaAlaLeuLeuAlaGluValAla 6798
      ||| ||| |||:|:|
1153 .....ATGCCGCTAGACATCTGCT 1173
      ||| ||| |||:|:|
6798 sPheLeuAlaArgGlyAlaLeuGlnProLeuProValGluIlePhePro 6814
      ||| ||| |||:|:|
1174 ACCCTGCTTTGGCGCATTTATCTGCGCGCATACCGAGCGCGCACG 1223
      ||| ||| |||:|:|
6815 LeuSerArgAlaAlaAspAlaPhe..ArgLysMetAlaGlnAlaGlnHis 6830
      ||| ||| |||:|:|
1224 ATTGGGTGCTGTGGAATTTGACGACGACGCTGCTTGTGACGTTGCG 1273
      ||| ||| |||:|:|
6831 LeuGlyLysLeuValLeuAlaLeuGlnAspProAsp..ValArgIleAr 6846
      ||| ||| |||:|:|
1274 TCTGCCCGGCGCA 1286
      ||| ||| |||:|:|
6846 yValProGlyGlu 6850
      ||| ||| |||:|:|

```

seq\_name: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:US-08-801-344-8

seq\_documentation\_block:

Sequence 8, Application US/08801344  
Patent No. 6087140

GENERAL INFORMATION:

APPLICANT: Cameron, Douglas C.

APPLICANT: Shaw, Anita J.

APPLICANT: Altaras, Nedim E.

TITLE OF INVENTION: MICROBIAL PRODUCTION OF 1,2-PROPANEDIOL

TITLE OF INVENTION: FROM SUGAR

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSER: Dewitt Ross & Stevens S.C.

STREET: 8000 Excelsior Drive, Suite 401

CITY: Madison

STATE: WI

COUNTRY: U.S.A.

ZIP: 53717-1914

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/801,344

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Sara, Charles S.

REFERENCE/DOCKET NUMBER: 09820, 037

TELECOMMUNICATION INFORMATION:

TELEPHONE: 608-831-2106

TELEFAX: 608-831-2100

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 510 amino acids

TYPE: amino acid

STRANDEDNESS: single

```

;
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Pyridine nucleotide transhydrogenase, subunit
; ORGANISM: A
US-08-801-344-8
```

## alignment\_scores:

```

Quality: 93.00 Length: 484
Ratio: 0.441 Gaps: 20
Percent Similarity: 43.595 Percent Identity: 19.835
```

## alignment\_block:

```
US-09-303-518D-127 x US-08-801-344-8
```

```
Align seg 1/1 to: US-08-801-344-8 from: 1 to: 510
```

```

100 GAAGATATGCGGTATGCGCCCTGATGAAAGTCAAGAGCGGATGC 149
    : : : : : : : : : : : : : : : : : : : : : : : :
46 Aspaaplysalaphevalglalaglyalagluilevalgluglyasnse 62
    : : : : : : : : : : : : : : : : : : : : : : : :
150 CGTCAAAAAGGCCAAGTGTGTTGAAGACAAAAGNATCCGGCGGTGG 199
    : : : : : : : : : : : : : : : : : : : : : : : :
62 rValtrpInserguileileuyls..... 71

200 TGTTCACGCGCCNGTTTCAGGCAAAATCGCCCATCATCGCGCC... 246
    : : : : : : : : : : : : : : : : : : : : : : : :
72 ..ValasnlalproleuaspaspulilealaleuasnproglYthr 87
    : : : : : : : : : : : : : : : : : : : : : : : :
246 ..... 246

88 ThrleuValserPheiletrproalaglinsnprogluleuemetglu 104
    : : : : : : : : : : : : : : : : : : : : : : : :
247 .....GAAAGCGCGTACTTTCAGTGCCTGATTCGCGTGAAGGCA 289
    : : : : : : : : : : : : : : : : : : : : : : : :
104 sleuAlaglulargasnVal.....ThValmetAlametsperserv 118
    : : : : : : : : : : : : : : : : : : : : : : : :
290 ACGAGCAATGCGAGTTCGACGCTAC.....GGCCCGCAAGCGTTGGCA 333
    : : : : : : : : : : : : : : : : : : : : : : : :
118 alproaglliserargAlaglnserleuaspAlaleuSerSermetAla 134
    : : : : : : : : : : : : : : : : : : : : : : : :
334 AACTTAAGCGCGGANGAANTNNGNCAATCTGATCCAA..... 372
    : : : : : : : : : : : : : : : : : : : : : : : :
135 AsnilealaglYtrargAlailevalglualalaahlsnslupheglYar 151
    : : : : : : : : : : : : : : : : : : : : : : : :
373 .....TCGCGTTTGTGACTGCGTGCCTGACCCGTCCTGACGAAA 415
    : : : : : : : : : : : : : : : : : : : : : : : :
151 gPhepserhethclYlnlletthralaAlaglylYvalproAlalaysv 168
    : : : : : : : : : : : : : : : : : : : : : : : :
416 TCCCTGCGCTCGAT.....GCC 432
    : : : : : : : : : : : : : : : : : : : : : : : :
168 almetValileglYalaglyalalaglyleuAlalaileglYalala 184
    : : : : : : : : : : : : : : : : : : : : : : : :
433 GAGCGCTTGGCCATCTTCGTCATGCGATGACCAATCCGCTGCGGC 482
    : : : : : : : : : : : : : : : : : : : : : : : :
185 AsmerleuaglYalalevalargAlapheaspthrargprogluvalY 201
    : : : : : : : : : : : : : : : : : : : : : : : :
483 AGACCGCTGTG.....GTTGTATCAAG 505
    : : : : : : : : : : : : : : : : : : : : : : : :
201 sglInlvalaglnsermetglYalaglupheleuagluleuaspPhe 218
    : : : : : : : : : : : : : : : : : : : : : : : :
506 AAGCGCGANGATTCAGACAGANGTNGCTATTCAGCCGCTTG... 552
    : : : : : : : : : : : : : : : : : : : : : : : :
218 lugalalaglyserglYasplYtrAlalalYvalmetSeraspAlaphe 234
    : : : : : : : : : : : : : : : : : : : : : : : :
553 .....ACGAGCGTAATAATTCATGTGTGTGAAGACGTCGCGCAGAC... 594
    : : : : : : : : : : : : : : : : : : : : : : : :
235 lIelYsalaglmetgluleuPhealalaAlaglnalalYsgluvalasp 251
```

```

595 .....GTGCGTGTGAAATGCTGCCACATGGA 625
251 eilevalThrThralaleuileProglYlYsProAlaprolYsleuileT 268
    : : : : : : : : : : : : : : : : : : : : : : : :
626 CACATGAATTCGGCGCGCCGATCCGCGGT..... 657
    : : : : : : : : : : : : : : : : : : : : : : : :
268 hrarglmetValaspsrmetYsalaglyserValilevalaspLeu 284
    : : : : : : : : : : : : : : : : : : : : : : : :
658 .....TTGAGTGGCAGCAGCATTCATTCATTGAGCC.....GT 692
    : : : : : : : : : : : : : : : : : : : : : : : :
285 AlAlaglnasnlglylYasnCysgluYtrThrValproglYgluleph 301
    : : : : : : : : : : : : : : : : : : : : : : : :
693 CGGTGCAACAAAACCGTTTGACCAATTCATTCAGATGTAATTCGCA 742
    : : : : : : : : : : : : : : : : : : : : : : : :
301 eThrThrglnasnlglylYalYsalileglYtrThraspleu..... 315
    : : : : : : : : : : : : : : : : : : : : : : : :
743 TCGGACGTTTGTTCGACACAGCGCTGTGAAACCGGCGGTGATGCT 792
    : : : : : : : : : : : : : : : : : : : : : : : :
316 .....ProglYargleuproThrglnserSerlnleu 326
    : : : : : : : : : : : : : : : : : : : : : : : :
793 TTGGGTGTCTTCAGTCAACAAACACGCGCTTTCGTCACGTTTGGG 842
    : : : : : : : : : : : : : : : : : : : : : : : :
327 TyrclYthrasnleuvalasn.....leuileuYleuclYslY 340
    : : : : : : : : : : : : : : : : : : : : : : : :
843 TCGCAAGATTCGCAATTCATTCGCGCGCAATGTGTGACGACACACC 892
    : : : : : : : : : : : : : : : : : : : : : : : :
340 sglulYsaspGlyAsnlethr.....ValaspleaspsY 353
    : : : : : : : : : : : : : : : : : : : : : : : :
893 GCGTATTCGCGT.....TCGATTCAGCGCGCGATTCACAGACGC 936
    : : : : : : : : : : : : : : : : : : : : : : : :
353 alValilearglyValThrvalileargAlaglylulethrtrpro 369
    : : : : : : : : : : : : : : : : : : : : : : : :
937 GCGCAGATTAATTTGGAGCTACCAATTCAGATTCGCTTATTCGACA 986
    : : : : : : : : : : : : : : : : : : : : : : : :
370 AlaProPole.....GlnValserAlaglProgl 380
    : : : : : : : : : : : : : : : : : : : : : : : :
987 AGCGCGCAGCAAAAG..... 1002
    : : : : : : : : : : : : : : : : : : : : : : : :
380 nAlaAlaglnlyAlaAlaProgluValYlYthrglugluYslYtrnc 397
    : : : : : : : : : : : : : : : : : : : : : : : :
1003 .....CTGTTTC 1008
    : : : : : : : : : : : : : : : : : : : : : : : :
397 YserProtrparglYsYtrAlaleuMetAlaleuAlaileleleuphe 413
    : : : : : : : : : : : : : : : : : : : : : : : :
1009 GGTGGGTGGCGCGCGAGCGGCAAAATCTGCATCAGCGGACAGACC.. 1056
    : : : : : : : : : : : : : : : : : : : : : : : :
414 GlyTrpMetAlaserValAlaProlysgluPheleuGlyHisPheThVa 430
    : : : : : : : : : : : : : : : : : : : : : : : :
1057 .....CTGCGCC 1063
    : : : : : : : : : : : : : : : : : : : : : : : :
430 lPheAlaleuAlagYsalValaglyTYtrYtrValValtrpsanValserH 447
    : : : : : : : : : : : : : : : : : : : : : : : :
1064 ATTTCTGAAAAACAACCTTCACAGTTCACGACAGCGGTCAACGGT... 1110
    : : : : : : : : : : : : : : : : : : : : : : : :
447 lAlaleuHisThrProleuMetSerValThrasnAlaileSerclYle 463
    : : : : : : : : : : : : : : : : : : : : : : : :
1111 .....GGCGACCGCGCCATGCGTGTGCTTGTATTCAGACGCGGT 1151
    : : : : : : : : : : : : : : : : : : : : : : : :
464 lIeValvalglY.....AlaleuLeuclnlelglynglylYtrpVa 478
    : : : : : : : : : : : : : : : : : : : : : : : :
1152 AATGCGCTAGACATCTGCTACCGCTGTTTGGCGGATTAATTCGTGC 1201
    : : : : : : : : : : : : : : : : : : : : : : : :
478 lserPheleuSerPheileAlaValleuileAlaserlIeasnlePheg 495
    : : : : : : : : : : : : : : : : : : : : : : : :
1202 GC 1203
    : : : : : : : : : : : : : : : : : : : : : : : :
495 lY 495
```

seq\_name: /cgn2\_6/plodata/1/1aa/6B\_COMB.pep:US-09-498-599-8

seq\_documentation\_block:

; Sequence 8, Application US/09498599  
; Patent No. 6303352

```

GENERAL INFORMATION:
APPLICANT: Cameron, Douglas C.
APPLICANT: Shaw, Anita J.
APPLICANT: Altaras, Nedim E.
TITLE OF INVENTION: MICROBIAL PRODUCTION OF
TITLE OF INVENTION: 1,2-PROPANEDIOL FROM SUGAR
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Demilt Ross & Stevens S.C.
STREET: 8000 Excelstor Drive, Suite 401
CITY: Madison
STATE: WI
COUNTRY: U.S.A.
ZIP: 53717-1914
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/498,599
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Sara, Charles S.
REFERENCE/DOCKET NUMBER: 09820.037
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-831-2100
TELEFAX: 608-831-2106
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 510 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHEITICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Pyridine nucleotide transhydrogenase, subunit
US-09-498-599-8

alignment_scores:
Quality: 93.00 Length: 484
Ratio: 0.441 Gaps: 20
Percent Similarity: 43.595 Percent Identity: 19.835

alignment_block:
US-09-303-518d-127 x US-09-498-599-8 ..

Align seg 1/1 to: US-09-498-599-8 from: 1 to: 510
100 GAAGAATATGCGCGTATGCGCCCTGATGATGAAGTCAAGAGGAGCATGCG 149
46 ASPPSPYSLALAPHEVALGIALAGLYALGLIILEVALGLULYSNSEN 62
200 TGTATTACCGCGCCGTTTCAGGCAAAATGCGCCGATCCATCGCGGCG... 246
62 RVALTRPGLNSERGLIILEULEULYS..... 71
150 CGTCAAAAAGGCCAAGTCTGTTTGAAGCAAAAAGNATCGGCGGTGG 199
72 VALASNALAPROLEUASPSAPGLIILEALAULEUASNPROGLYTHR 87
246 ..... 246
88 ThrleuValserPheIleTrrProAlaGlnAsnProGluLeuMetGlnly 104
247 ..... GAAGAAGCGCTACTTCAAGTCGTCGTCGATTCGCGTTGAAGCA 269
|||||.....

```

```

104 sLeuAlaGluArgAsnVal.....ThrValMetAlaMetAspSerV 118
290 AGACGAATTCAGATTGCAACGCTAC.....GGCGCCGAGCGTTGGCA 333
118 aProArgIleSerArgAlaGlnSerLeuAspAlaLeuSerMetAla 134
334 AACTTAAGCGCGCAGNANTMNGNNGCAATCTGATCCAA..... 372
135 AsnIleAlaGlyTyrArgAlaIleValGluAlaIleGluPheGlyAr 151
373 .....TCCGGTTTGTGGACCTGCGCTGATTCAGCGTTTCAGCAAA 415
151 gphePheThrGlyGlnIleThrAlaAlaGlyLysValProProAlaLysV 168
416 TCCCTGCGCGTCGAT.....GCC 432
168 alMetValIleGlyAlaGlyValAlaGlyLeuAlaIleGlyAlaIle 184
433 GAGCGGTTTCGCATCTTCGTAATGCGATGACACCAATCCGCTGCGGC 482
185 AsnSerLeuGlyAlaIleValArgAlaPheAspThrArgProGluVally 201
483 AGACCTCTG.....GTTGATCAAG 505
201 sGluGlnValGlnSerMetGlyAlaGluPheLeuGluLeuAspPheLys 218
506 AAGCGCGNAGATTCAGACGANGTGTGCTGATTGACCGCTTG... 552
218 IuGluAlaGlySerGlyAspGlyTyrAlaLysValMetSerAspAlaPhe 234
553 .....ACCGAGCTAAATCCATGCTGTGAAGCACTGGCCGACG... 594
235 IleLysAlaGluMetGluLeuPheAlaIleAlaLysGluValAspII 251
595 .....GTCCGCTGAAATGCTGCAACATCGAA 625
251 eIleValThrThrAlaLeuIleProGlyLysProAlaProLysLeuIleT 268
626 CACATGAATTCGCGCGCCGCGATCCGCGGT..... 657
268 hrArgGluMetValAspSerMetLysAlaGlySerValIleValAspLeu 284
658 .....TTGAGTGCACGCGACATTCATTCTTACGCG.....GT 692
285 AlaAlaGlnAsnGlyLysAsnGlyLysThrValProGlyGluIlePh 301
693 CGGTGCAAAACAAACCGTTTGACATCAATTAATCAAGATGTAATGCCA 742
301 eThrThrGluAsnGlyValLysValIleGlyTyrThrAspLeu..... 315
743 TCGGACGTTTGTTCGAACAGCGCGCTGAAACACCGAGCGCGTATGCT 792
316 .....ProGlyArgLeuProThrGlnSerSerGlnLeu 326
793 TTGGTGTTCTCAAGTCAACAAACGCGCTCTGGCTACCGTTTGGG 842
327 TyrGlyThrAsnLeuValAsn.....LeuLeuLysLeuLeuCysLy 340
843 TCGGAAGATGCAAAATTAATTCGCGGCAATGTTGTTGACGCAACAAC 892
340 sGluLysAspGlyAsnIleThr.....ValAspPheAspAspV 353
893 GCGTGAATTCGCGT.....TCGGTATTGAACGCGCGCATACACAAGC 936
353 alValIleArgGlyValThrValIleArgAlaGlyTgtGltTrrPro 369
937 GCGCAGATTATTGGACGCTACACCAATCAAGATTCGTTACGATGAGA 986
370 AlaProPhe.....GlnValSerAlaGlnProGlu 380
987 AGCGCGCACAACAAG..... 1002
380 naIaAlaGlnLysAlaAlaProGluValLysThrGluGluLysCysThrC 397

```



```

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/783,774
FILING DATE: 15-JAN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7682-037
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 907 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-783-774-2

```

```

Alignment scores:
Quality: 91.00 Length: 388
Ratio: 0.569 Gaps: 17
Percent Similarity: 41.237 Percent Identity: 21.134

```

alignment block:

US-09-303-518d-127/rev x US-08-783-774-2 ..

Align seg 1/1 to: US-08-783-774-2 from: 1 to: 907

```

1133 ATGCGACCATGAGCGCGTCCACCGTTGAGC.....GC 1099
      |||||
396 LeuGlyThrAlaProLysThrLeuIleIleThrArgThrAlaThrAsnAl 412
1098 TGTCGTGAACCTGAGAGTTGTTTTCAGAAATGCGGAGGTCGTAC 1049
      |||||
412 aThrThrThrThrHisLysValIlePheSerLysAlaProGlySerThrT 429
1048 GCCTGATGAGATTTGTCGGGTGGGGGCAACCCAGCGAACACAGCTCT 999
      |||||
429 hThrSerProThrLeuAsnThrGlyPheAlaSerProAsnThrThr 445
998 TTGCTCGCGCCTTTCGATACGGAATCTGATGTGAGCTCCCAA 949
      |||||
446 ThrGlyLeuProSerSer..ThrHisVal..... 454
948 ATAAATCGCGCGCTTGTTGTAATCGCGCGCTCATACGGAACGGAAA 899
      |||||
455 .....ProThrAsnLeuThrAlaProAlaSerThrGlyProThrTy 468
898 TCACGGCGGTGTCTGCGTCAACCAATTCGCGCGCAATATTCGCACTAC 849
      |||||
468 aLysThrAlaAspValThrSerProThrProAlaGlyThrThrSerGly 484
848 TTCGCACCCCAAAACGGTAGCGAAGAG.....CGTGGTTTGTTCAC 808
      |||||
485 AlasThrProValThrProSerProSerProThrPaspAsnGlyThrGlu 501
807 TTGAGACCAACCCCAAGCATTCACGCGCGGTGTTCAGACGGCGCTGTG 758
      |||||
501 rLysAlaProAspMetThrSerSerThrSer.....ProValT 514
757 CAACCAAAACGTCCGATGCAATTCATCTGTATTAATGATGTCACAAAG 708
      |||||
514 hThrThrProThrProAsnAlaThr..... 521
707 GTTTGTTGTCACCGACCGGCTCAATGAATGAATGTCGCTGCACCTCAA 658
      |||||

```

```

522 .....SerProThr..... 524
657 ACCGGCGGATGGGGCGCCGGAATTCATGTTGATGTTGCAGCAT 608
      |||||
525 .ProAlaValThrThrProThrProAsnAlaThrSerProThrProAla 541
607 TTTCAGACGGCAGCTCTGCGCGAGTCCCTTACACATGATGATTTACGC 558
      |||||
541 aLthr.....ThrProThrProAsnAla..... 548
557 TCGGTCAACGGCTCAATACCAACCAATCGCTCAATATCNCNGCGGC 508
      |||||
549 .....ThrSerProThrLeuGlyLysThrSerProThr 559
507 TTCTTTGATCAACAC...ACAGGCTCTGCGCGCAGCATTTGTC.... 465
      |||||
559 rSerAlaValThrThrProThrProAsnAlaThrSerProThrLeuGly 576
464 .....TCCATGCAATTGACGAAGATGGCGAAGCGCTGGCA 429
      |||||
576 yThrSerProThrSerAlaValThrThrProThrProAsnAlaThrSer 592
428 TCGACGGCAGGATTTTGTGACGAGCGAGTACGACGCGCATCCACAA 379
      |||||
593 ProThrLeuGly..... 596
378 ACCGGATTGATCAGATTGCNCNANNANTTCGCGCTTAAGTTTGCCA 329
596 ..... 596
328 ACGCTTCGGCGCGGTAGCGTTGCACTGATTCGTGCTGCTTCACAG 279
      |||||
597 .....LysThrSerProThrSer 602
278 GCATTCAGACGACGACTGAGTACGCGCTTTCGCGCGATGATGAGCGGC 229
      |||||
603 AlaValThrThrProThrProAsnAlaThrGlyProThr..ValGlyG 618
228 GATTTGCTGAAACNGCGCGGTAAACACCAACCGCGGATNCTTTTGT 179
      |||||
618 uThrSerProGlnAlaAsnAlaThrAsnHisThr.....LeuG 631
178 CTTCAACAGACATTTGGCCTTTTTCAGCGGATGCGCTTCTGTACTTC 129
      |||||
631 LylGlyThrSerProThrProValAlaThr..... 640
128 ATCNAAGGGCGCATACCGGCAATATCTCGCAAGCAACGCACTTGC.. 81
      |||||
641 .....SerGlnProLysAsnAlaThrSerAl 649
80 .GTAAATGACGGCGCGCTCATAAATGACTGCTCGCGGTGCGCGCGATG 32
      |||||
649 aValThrThrGlyGlnHisAsnIleThrSerSerSerThrSerSerMet 666
31 CGAGCTTTAGACT 18
      |||||
666 er....LeuArgPro 669

```

seq\_name: /cgn2\_6/ptodata/1/lae/PCTUS\_COMB.pep.PCT-US95-04611A-19

```

seq_documentation_block:
; Sequence ID, Application PC/TUS9504611A
; GENERAL INFORMATION:
; APPLICANT: Spaele, Richard and Jackman, Winthrop, T.
; TITLE OF INVENTION: Non Splicing Variants of gp350/220
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESS: Cooley Goddard Castro Huddleson & Tatum
; STREET: 5 Palo Alto Square
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94306

```



COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04611A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/229,291  
FILING DATE: April 18, 1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Luann Cserf  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: AVIR-003/00US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-843-5163  
TELEFAX: 415-857-0663  
TELEX: 380816 COLEAPPA  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 907 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-04611A-19

alignment\_scores:  
Quality: 91.00 Length: 388  
Ratio: 0.569 Gaps: 17  
Percent Similarity: 41.237 Percent Identity: 21.134

alignment\_block:  
US-09-303-518D-127/rev x PCT-US95-04611A-19 ..  
Align seg 1/1 to: PCT-US95-04611A-19 from: 1 to: 907

1133 ATGGAGCCAGTGGGGGCGGCCAGCCCTTGAC.....GC 1099  
396 LeuLylThrAlaProLysThrLeuIleIleThrArgThrAlaThrAsnAl 412  
1098 TGTGTCGTAAGTGAAGAGTTGTTTTCAGAAATGCGGAGGCTGCTAC 1049  
412 aThrThrThrHisIleValIlePheSerLysAlaProGluSerThrT 429  
1048 GCGGATGAGTATTTGTCGCGGCGGCGCAACCCGAGCAACAGCTCT 999  
429 hThrSerProThrLeuAsnThrThrGlyPheAlaAspProAsnThrThr 445  
998 TTGCTGGGCGCTTCGATTAAGGAATCTGATTGTGAGCTGCCAA 949  
446 ThrGlyLeuProSerSer...ThrHisVal..... 454  
948 ATATGTCGGCGCTTGTGTATCGCGCGTTCATACCAAGCAAGGANA 899  
455 .....ProThrAsnLeuThrAlaProAlaSerThrGlyProThrV 468  
898 TCACGGGCTTCTGCGTCAACCAATTCGCGCAAGTAAATTCGATACT 849  
468 aIleThrAlaAspValThrSerProThrProAlaGlyThrThrSerGly 484  
848 TTCGACCCCAAAAGGTAGCAAGAG.....CGGCTTTTGTGAC 808  
485 AlaSerProValThrProSerProSerProThrPaspAsnGlyThrGluSe 501  
807 TTGAGAACCCCAAGCAATCAGCGCTGCTGTTCAAGAGCGCTGTG 758  
501 rLysAlaProAspMetThrSerSerThrSer.....ProValT 514  
757 CAAGCAAAAGCTCGATGGCATTTACATCTTGATAATTGATGATCAAG 708

514 hThrProThrProAsnAlaThr..... 521  
707 GTTTTGTTCACCGACCGGCTCAATGAATGAATGTCGTCGACTCA 658  
522 .....SerProThr..... 524  
657 ACCGGCGGATGGGCGCCGCAATTCATGTCTTGCATGTGGCAGCAT 608  
525 .ProAlaValThrThrProThrProAsnAlaThrSerProThrProAlaV 541  
607 TTTCAGACGGCAGCTGCGCGAGTCCCTACACACATGATTTTACG 558  
541 alThr.....ThrProThrProAsnAla..... 548  
557 TCGGTAAAGCGCTCATACCAAGCANACNTGCTGAATCTGCGGCGC 508  
549 .....ThrSerProThrLeuGlyLysThrSerProTh 559  
507 TTCTTGATCAACAC...ACAGGCTGCGCCGCGGCGGATTTGTC... 465  
559 rSerAlaValThrThrProThrProAsnAlaThrSerProThrLeuGlyL 576  
464 .....TCCATGCGATTGACGAGATGCGCAACGCGCTGCGCA 429  
576 yThrSerProThrSerAlaValThrThrProThrProAsnAlaThrSer 592  
428 TCGACGCGCAGGATTTTGTCTGAACGAGCGGATACGACGCGAGTCCACA 379  
593 ProThrLeuGly..... 596  
378 ACCGATTTGATCAGATTGCGNCCNANNTTCNTGCGCGTTAAGTTGCCA 329  
596 ..... 596  
328 AGCTTCGCGCGCGGAGCGTTGCACTGATTCGCTGCTGCTGCTCAAG 279  
597 .....LysThrSerProThrSer 602  
278 GCAATCAGACCGACTGAAGTACGCGCTTTGCGCGGATGATGAGCGGC 229  
603 AlaValThrThrProThrProAsnAlaThrGlyProThr...ValGlyG 618  
228 GATTTTGCCTGAACGCGCGGATGAACACACACCGCGGATTCCTTTGT 179  
618 uThrSerProGlnAlaAsnAlaThrAsnHisThr.....LeuG 631  
178 CTGCAACAGCACTTGCGCTTTTGTGACGCGATGCGCTTCCTGACTTTC 129  
631 LysGlyThrSerProThrProValThr..... 640  
128 ATCNAGGGGCGCATACCGGCTATTTCTTGGCAGCAAGCGACTTGC.. 81  
641 .....SerGlnProLysAsnAlaThrSerAl 649  
80 .GTAATGAGGCGCGCATTAATGACTGCTCGGCTGCGCGGATG 32  
649 aValThrThrGlyLlnHisAsnIleThrSerSerSerThrSerSerSer 666  
31 GCAGGTTTAGACCT 18  
666 er...LeuArgPro 669

seq\_name: /cgn2\_6/plodata/1/1aa/5A\_COMB.pep:US-08-127-499A-26  
seq\_documentation\_block:  
; Sequence 26: Application US/08127499A  
; Patent No. 5510264  
; GENERAL INFORMATION:  
; APPLICANT: VAN ALSTYNE, Diane  
; APPLICANT: SHARMA, Lawrence Rajendra  
; TITLE OF INVENTION: ANTIBODIES WHICH BIND MENINGITIS RELATED  
; TITLE OF INVENTION: HOMOLOGOUS ANTIGENIC SEQUENCES



## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/482,847  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 514  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/127,499  
 FILING DATE: 28-SEP-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BENT, Stephen A.  
 REGISTRATION NUMBER: 29,768  
 REFERENCE/DOCKET NUMBER: 51916/104/INBI  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202)672-5300  
 TELEFAX: (202)672-5399  
 TELEX: 904136  
 INFORMATION FOR SEQ ID NO: 26:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 484 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: unknown  
 US-08-482-847-26

alignment\_scores:  
 Quality: 88.00 Length: 347  
 Ratio: 0.518 Gaps: 17  
 Percent Similarity: 48.991 Percent Identity: 20.173

## alignment block:

US-09-303-518d-127 x US-08-482-847-26 ..

Align seg 1/1 to: US-08-482-847-26 from: 1 to: 484

```

145 GATGCCGTCAAAAAGGCCAAGTCGTGTTGAAGACAAAAGATCCGGG 194
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
50 AAspAlaIleLysLysAlaAsnAsnLeuThrThrAspLysIleValPro1 66
195 CGTGGTGTTCACGCGCGCNGTTCAGAGCAAAATCCGCCCATCCAGCGG 244
-|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
66 yGln.....LysLeuGlnValAsnAsnGluValAlaIleAlaGluLysT 81
245 GCGAAAGCGCGTACTT..... 261
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
81 hrcGlnLysSerValSerAlaThrTrpLeuAsnValArgSerGlyAlaGly 97
262 .....CAGTCGCGTGTGATTCGCCGTGTAAGCAGCAGCAATCGAGTT 305
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
98 ValAspAsnSerIleIleThrSerIleLysGlyThrLysValThrVa 114
306 CGAAGCGTACGGCCCGAAGCGGTGGCAAACTTAAGCGCGCANGANTNN 355
|||:|||||:|||||:|||||:|||||:|||||:|||||:
114 LGu..... 115
356 GNNCGCATGTGATCCAAATCCGTTGTGGACTGCGTCGTANCCGTCG 405
116 .....ThrThrIuSerAsnGlyTrp..... 122
406 TTCACCAAAATCCCTGCGCGTGCAGCGCAGCGCGTCCGCAATCTGTCAA 455
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
123 ..HisLysIleThrLysAsnAspLys.....ThrGlyPheValAs 136
456 TCGGATGAGC...ACCAATCCGCTGCGCGCAGACCCCTGTGTTGATGATCA 502
|||:|||||:|||||:|||||:|||||:|||||:|||||:
136 nelyLysTyLeuThrAspLysAlaValSerThrProValAlaProThrG 153
503 AAGAGCCCGCANGATTTTCAGACGANGTNTGCTGGTATTTGAGCGCTTGG 552
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

153 InGUValLysLysGlu.....ThrThrGlnGlnAlaAlaProAla 167
553 ACCGAGCGTAAATCCATGTCGTAGCAGCGTGGCGAGCGAGTGGCGC 602
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
168 AlaGluThrLysThrLysGluValLysGlnThrGlnAlaThrThrProAl 184
603 TGAATAATGCTGCCACATCAATGAAACATGAATTCGGCGCCGATCCGG 652
-|:|||||:|||||:|||||:|||||:|||||:|||||:
184 aProLysValAla.....GluThrLysGlu.....ThrProValAla 197
653 CCGGTTTGAGTGGCAGCGACATTCATTTTCATTGAGCCGCTCGCTCAAC 702
|||:|||||:|||||:|||||:|||||:|||||:|||||:
197 spcLysAsnAlaThrThrHis.....AlaValLysSerGly 208
703 AAACCGTTTGGACCATTCATTCAGATGTAATTCATTCATTCAGCGATT 752
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
209 AspThrIleTrpAlaLeuSerValLysTyGlyValSerValGlnAspI 225
753 GTTTGCAACAGCCCGCTGACACCGAGCGCGTATTCCT.....T 793
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
225 emetSerTrpAsnAsnLeuSerSerSerIleTyValGlyGlnLysL 242
794 TGGGTGTTCTCAGTCAACAAACACCGCTCTTTC..... 829
|||:|||||:|||||:|||||:|||||:|||||:|||||:
242 euAlaIleLysGlnThrAlaAsnThrAlaThrProLysAlaGluValL 258
830 .....GTACCGTTTGGTGGCGAAGTATCCGAAATTCATTCAGCGCGCA 872
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
258 sThrGlnAlaProAlaIleGlnLysGlnAlaAlaProValLysGlu 275
873 ATTGGTTCAGCAGACCAACCGCGTATTCGCGTTCGATTCGATTCAGCGG 922
|||:|||||:|||||:|||||:|||||:|||||:|||||:
275 sn.....ThrAsnThrAsn.....ThrAla 281
923 CGATTACACAGCGCGCAGCATTTATTGGACCGCTACCAATCAATCAATT 972
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
282 ThrThrGlnLysLysGlnThrAlaThrGlnGlnInLThrAlaProLysAl 298
973 TCCG.....TTATCGAAGAGCGCGCAGCAAAAGCTGTTCGGCTGGGT 1016
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 aProThrGlnAlaLysLysProAlaProAlaProSerThrAsnThrAsn 315
1017 TGGCGCGCAGCGCGCAAAATCTCCATCAGCGCGTACGACCCCTGGCCATT 1066
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
315 LaAsnLysThrAsnThrAsnThrAsnThrAsnThr..... 326
1067 TCCTGAAAACAACTCTTCAGTTCACGACAGCGCGTCA 1105
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
327 .....AsnThrAsnAsnThrAsnThrAsnThrProSer 337

```

seq\_name: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:US-08-907-166-10

## seq\_documentation block:

Sequence 10, Application US/08907166  
 Patent No. 5948666  
 GENERAL INFORMATION:  
 APPLICANT: Callen, Walter  
 APPLICANT: Mather, Eric  
 TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF NOVEL POLYMERASES  
 FILE REFERENCE: 09010/027001  
 CURRENT APPLICATION NUMBER: US/08/907,166  
 NUMBER OF SEQ ID NOS: 12  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 10  
 LENGTH: 762  
 TYPE: PRT  
 ORGANISM: Desulfurococcus sp.  
 FEATURE:  
 NAME/KEY: variation  
 LOCATION: (601)..(601)  
 OTHER INFORMATION: Xaa at position 601 is alanine or proline  
 US-08-907-166-10

alignment\_scores:      Quality:    87.50      Length:    400  
                          Ratio:      0.494      Gaps:      15  
                          Percent Similarity: 44.250      Percent Identity: 19.750

alignment\_block:  
 US-09-303-518D-127 x US-08-907-166-10 ..

Align seg 1/1 to: US-08-907-166-10 from: 1 to: 762

```

169 CTGTTTGACGACAAAGNATCCGGCGTGTGTTACCCGCCGCTTTC 218
      |||
39  IletyraspaasnArgleuileProglymetArglyValThrAspPhe 55
      |||
219 AGGCAAAATCGCGCCATCCGCGGCAAAAGCGCTTCTACTCGG 268
      |||
55  rasnGlylaGlnlysleuValmetValysProGluileProGlnserL 72
      |||
269 TCGTGATTGCCGTGAGCAACGACGAATCGATTGCAACGCTACGCG 318
      |||
72  euValGluValArg.....GluLeu...PheArgAsnGluPro 84
      |||
319 CCCGAAGCGTTGCGCAACTTAAGCGCGGANGANTNNGNCCATCTGAT 368
      |||
85  ProGluThrVal.....LysLeuAl 91
      |||
369 CCAATCCGTTTGTGACTGCGCTGCGTANCCGCTTCAGCAAAATCC 418
      |||
91  aGluGluLeuLeuLeuPheGluGluSerProProlyValysArgV 108
      |||
419 CTGCGCGTATCCGACGCCGCTTCCGCTTCTGCAATGCGATGACAC 468
      |||
108 aLaValaValspilleGluValPheThrProPheLysGlyArgileProser 124
      |||
469 AATCCGCTNGCGGACAGACCTGTGTTGATCAAGAAGCGCGANGA 518
      |||
125 ProLysLeuAlaGluTyrProValileSerileAlaLeuAlaGlySerAs 141
      |||
519 TTTCAAGACGANTTNGCTGTAATTGAGCCGTTTGACCGGTAATAATCC 568
      |||
141 pGlyLeuLysLysLleLeuLeuAlaArg.....GluTyrLys...H 155
      |||
569 ATGTGTGTAGAGCAGCTGGCGGACGCTGCG..... 600
      |||
155 IsAspPheAspTyrMetMetGluAspTyrProValGluAlaGluValGlu 171
      |||
601 .....TCTGAAATGCTGCGCAACATCAACACATGAATTCGGCGG 641
      |||
172 ValPheAspSerGluLysAspMetLeuLeuAlaPheArgileMetGle 188
      |||
642 CCGGCATCCGGC..... 654
      |||
188 ySerTyrProValValleuThrTyrAsnGlyAspAsnPheAspLeuGluT 205
      |||
655 .....GTTTGAGTGCACGACATTCAT 678
      |||
205 yLeuTyrValArgAlaPheLysLeuGlyLleLeuArgSerHisIle... 220
      |||
679 TTCATTGAGCGCGTGGTGCACAAACCGTTTGACCATCAATTAATCA 728
      |||
221 .....ProLeuLysLleGlyGluAspMetLleArgLileAspPhe 234
      |||
729 AGATGTAATTCGATCGAGCTTGTTCGACA..... 762
      |||
234 rIleHisLeuAspLeuTyrLysPhePheSerAsnArgAlaValLysAsnT 251
      |||
763 .....GGCGCTGACACCGGCGCGCTGATTGCTTGGTGGTGT 801
      |||
251 yralaPheGlyLysTyrGlnGlnGluLysLeuAspAlaValSerGly 267
      |||
802 TCT.....CAAGTCAACAAACACGCGCTTGTGCTACGCTTTGGGTGC 845
  
```

```

268 AlaleuLeuGlyValSerLysIleGlyPheGluGluThrIleGlyLyl 284
      |||
846 GAAGTATCCGCAATTCGCGGCGAATGTTGTCGCGACGACACCGCG 895
      |||
284 eSerAlaSerLeuLeuAlaValTyrAsnTyrArgAspAlaGlu..... 298
      |||
896 TGATTTCCGGTTCGATTGACGCGCGCATTCACAAAGCGCGCACGAT 945
      |||
299 .....IleThrLeuAsn..... 302
      |||
946 TATTTGGACGCTACCAATTCAGATT..... 972
      |||
303 ...LeuThrMetPheSerAsnGluLeuValTyrLysLeuIleIleLeu 318
      |||
973 .....TCCGTTATCGAAGAAGCGCGACGACGACGCTGT 1006
      |||
318 uAlaArgValSerLysThrSerLleGluAspLeuLysArgArgGlnIle 335
      |||
1007 TCGGCTGGGTTGCCGCGACGCGCAAAATCTCCATCAGCGCTACGACC 1056
      |||
335 eTyrTyrPhe.....GlnAsnLeuPhePheThrGluArgArgLys 348
      |||
1057 CTGCGCATTTCTGAAAAACAACTTTCAGTTCACGACCGCCGCA 1106
      |||
349 LeuGlyTyrLeuIleProAsnLysGluAspLleLeuArgHisValArg 365
      |||
1107 CGGTGCGACCGCGCATGCTG..... 1128
      |||
365 yThrGlyThrLysAlaIleIleGlyGlyLysTyrAlaGlyAlaLeu 382
      |||
1129 .....CCGATTGTACTTACGAGCGCGTAATGCGCGCTAGCATC 1167
      |||
382 aValaGluProProLysGlyAlaPhePheAsnValaValleuAspIle 388
      |||

```

seq\_name: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:US-09-351-200-2

seq\_documentation\_block:

```

; Sequence 2, Application US/09351200
; Patent No. 6320033
; GENERAL INFORMATION:
; APPLICANT: BOUDBONNAIS, Yves
; APPLICANT: LAMARRE, Claude
; TITLE OF INVENTION: CANDIDA ALBICANS GENE (CSA1) ENCODING A
; FILE REFERENCE: 6013-71"US" CC/
; CURRENT APPLICATION NUMBER: US/09/351,200
; EARLIER FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: CA2,237,134
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1203
; TYPE: PRT
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (0)...(0)
US-09-351-200-2

```

alignment\_scores:      Quality:    87.50      Length:    334  
                          Ratio:      0.534      Gaps:      10  
                          Percent Similarity: 49.102      Percent Identity: 20.958

alignment\_block:  
 US-09-303-518D-127/rev x US-09-351-200-2 ..

Align seg 1/1 to: US-09-351-200-2 from: 1 to: 1203

```

1108  SerSerAlaAlaSerValGlnLeuSerAlaAlaAlaAlaGlnLysSerSe 11224
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
468  GSTGTTCATTCGCATTGACGAGAGATGGCGAACGGCTCGCATCGACGGAG 419
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1124  rGlnAlaValGlnValAlaGlnThrAlaValAlaGlnAlaSerLysAlaG 1141
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
418  GG 417
      ||
1141  Ly 1141
      ||
seq_name: /cgn2_6/ptodata/1/1aa/PTMUS.COMB.depi:PT-US94-00198-4

```

```

seq_name: /cpg2_6/pdata/1/1aa/PCTUS_COMB.pep:PCT-US94-00198
seq_documentation_block:
? Sequence 4, Application PCT/US9400198
? GENERAL INFORMATION:
? APPLICANT: Schering Corp.
? TITLE OF INVENTION: RAS Associated GAP Proteins
? NUMBER OF SEQUENCES: 6
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Schering Corp.
? STREET: 1 Giraldo Farms
? CITY: Madison
? STATE: New Jersey
? COUNTRY: USA
? ZIP: 94304-1104
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: Macintosh
? OPERATING SYSTEM: 6.0.8
? SOFTWARE: Microsoft Word 5.1a
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: PCT/US94/00198
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/004,824
? FILING DATE: 15-JAN-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Lunn, Paul G.
? REGISTRATION NUMBER: 32,743
? REFERENCE/DOCKET NUMBER: DX0352 PCT
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (201)822-7255
? TELEFAX: (201)822-7039
? INFORMATION FOR SEQ ID NO: 4:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 3079 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? ORIGINAL SOURCE:
? ORGANISM: Saccharomyces cerevisiae
PCT-US94-00198-4

Alignment_scores:
? Quality: 87.50 Length: 149
? Ratio: 1.017 Gaps: 3
? Percent Similarity: 57.718 Percent Identity: 25.503

alignment_block:
US-09-303-518D-127/rev x PCT-US94-00198-4 ...

Align seg 1/1 to: PCT-US94-00198-4 from: 1 to: 3079

506 TCTTTGATCACACCCAGGGCTGCGCCGACGATGGTGCATCGC 457
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
387 SermettietlrrtlrlnasglnasrlahhlsclnlyserseerProSe 403
456 ATTGACGAAGATGGCGAAGGGCTGCGCATCGACGGCA..... 420
: : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : :

```







```
alignment_scores:
  Quality: 87.00      Length: 226
  Ratio: 0.784      Gaps: 6
Percent Similarity: 49.115      Percent Identity: 22.566
```

```
alignment_block:
ns-09-303-518D-137/rev x ns-09-323-433A-14
```

Align seg 1/1 to: US-09-323-433A-14 from: 1 to: 475

```

764 CCGTGGTCAAAACCAAGCTCCGATGAGCAATTAATCTGTAATGATGCT 711
      |||:::|||||::: ||| ||||| :::::
57 ProLalaserGluValAlaThrAlaAlaThrSer.....AlaLeI 71
      : || :::::|||||::: |||
714 CCAAAACGGTTTGTGGTCAACCGACCGGCTCAATGAATGATGCGTGC 665
      : || :::::|||||::: |||
71 eThrThrSerValLeuAlaThrProThrSer..... 81
      : || :::::|||||::: |||
664 CACTCAAAACCGCGGATGGCGGGCCCGCAATTCAGTGTGGATGTTG 615
      ||||| ::: ::: |||
82 .....AlaAlaAlaGlyLeuAlaAlaSerIleLeuAlaValSerAlaIa 96
      : ||| :::::|||||::: |||
614 GCACCATTTTCAGACGGCAGCTGTCGGCCCGATGCTTACACATGAGAT 565
      ||||| ::: |||
97 LeuAlaIysAsnGluIuIyIleSerAspAlaAla..... 108
      ||||| ::: |||
564 TTATCGCTCGGTCAAAAGCGCTCAATACACAGCAACATCGTGAATCMT 515
      ||||| ::: |||
109 .....SerAlaThrAlaSerThrSerg 116
      : ||| :::::|||||::: |||
514 CGACGGCTTCTTTATACAAACACAGGGTCTGCCGCAACGGATTTGGTG 465
      :: ||||| :::::|||||::: |||
116 IngIyAlaSerSerSerSerSerSerSerAlaThrSerThrLeuGlu 132
      : ||||| :::::|||||::: |||
464 TCATATCGATGATGAAGATAGCCAAAGCGCTCGGCGATCGACAGCAGGAT 415
      ||| :::::|||||::: |||
133 SerSerSerValSerSerSerSerGluAlaAlaThrProThrSerThrVa 149
      : ||| :::::|||||::: |||
414 TTTGCTGAACGAGCGGNTACGACGCGCATGCCAAACCGGATGTGATCA 365
      :::
149 Val..... 150
      : ||| :::::|||||::: |||
364 GATTTCNNCANNANTTCNTGCCCGCTTAAGTTTGGCAACGCTTGGGGGCG 315
      : ||| :::::|||||::: |||
151 .....SerThrSerSerAlaThrGlnSerSerAlaSerSerAla 165
      : ||| :::::|||||::: |||
314 TAGCGTTCAACTGCATTTCTGCTGCTGCTCAACAGCAATCACAGCGA 265
      : ||| :::::|||||::: |||
164 ThrIySerSerThrSerSerThrSerProSerThrSerThrSerThrSe 180
      : ||| :::::|||||::: |||
264 CTGAAGTACGCGCTTTTCGCGCGATGATGGCGGCAATTTGGCTGAAA 215
      : ||| ::: |||
180 ThrSerSerThrSerSer.....Ser 186
      : ||| :::::|||||::: |||
214 CAGCGGGCGGTAAACACAGCGCGGATTCCTTTGTCTTCAACAGCACT 165
      : ||| :::::|||||::: |||
188 eSerSerSerSerSerSerSerSerSerSerSerSerGlySerIle 205
      : ||| :::::|||||::: |||
164 TGGCGTTTTCAGC.....GCATGCGCTTCCTTGACTTTTCATCMAGG 125
      : ||| :::::|||||::: |||
120 GCGCATACCGGCAATTTCTGCGCAAC 93
      : ||||| ::: |||
221 yThrIleProCysAspIySphProSer 230
      : ||||| ::: |||

```

seq\_name: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:US-08-967-513-5

```
seq_documentation_block:
; Sequence 5, Application US/08967513
; Patent NO. 5783436
; GENERAL INFORMATION:
; APPLICANT: Robert P. Hausinger
```

```

TITLE OF INVENTION: Mutant Urease and Method
TITLE OF INVENTION: of Use For Determination
TITLE OF INVENTION: of Urea
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ian C. McLeod
STREET: 2190 Commons Parkway
CITY: Okemos
STATE: Michigan
COUNTRY: USA
ZIP: 48864
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 Inch, 360 Kb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,513
FILING DATE: 11-NOV-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/687,645
FILING DATE: July 26, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Ian C. McLeod
REGISTRATION NUMBER: 20,931
REFERENCE/DOCKET NUMBER: MSU 4.1-309
TELECOMMUNICATION INFORMATION:
TELEPHONE: (517) 347-4100
TELEFAX: (517) 347-4103
TELEX: No. 5783436e
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 567
TYPE: amino acids
STRANDEDNESS: Single
TOPOLOGY: linear
MOLECULE TYPE:
DESCRIPTION: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: N/A
STRAIN: N/A
INDIVIDUAL ISOLATE: N/A
CELL TYPE: N/A
FEATURE:
NAME/KEY: subunit Urec
LOCATION:
IDENTIFICATION METHOD: Sequencing
OTHER INFORMATION: Encoded subunit of mutant
OTHER INFORMATION: urease
US-08-967-513-5

```

```

alignment_scores:      86.50      Length:      247
                       Quality:    0.832      Gaps:      12
                       Ratio:      42.105     Percent Identity: 23.077
Percent Similarity:
alignment_block:
US-09-303-518D-127 x US-08-967-513-5      ..
Align seg 1/1 to: US-08-967-513-5 from: 1 to: 567

```

541 TTAGAGCGTTTACACGAGGTCAATTCATGTGTGTAGGACGTCGCG 580  
 260 Deu1a1a1a1leclgylgrgthrlfhnsthrfhnsthrfnglucylal 276  
 591 AGAGCTGCGCTGTGAATAATGCTGCCACATCGAATCAATCAATTCGCG 640  
 276 a.....dlg 278

```

641 GCCCGCATCCGGCCGTTTGAGTGGCAGCACATTTCAT...TTCATTGAG 687
    |||
278 LYGlyHisAlaProAspIleThrIleAlaHisProAsnIleLeu 294
    |||
688 CCGGTCGGTGCACAAACCGCTT...TGGACCATCAATTATGACAGATGT 734
    |||
295 ProSerSerThrAsnProThrLeuProTyrThrLeuAsnThrIleAspG 311
735 A.....ATTGCCA 742
311 whisLeuaspMetLeuMetValCysHisHisLeuAspProAspIleAla 328
743 TCGGACGCTTTGTTGCAACAGCCGCTGTGAACACCGCCGTTGATTCGT 792
    |||
328 LuAspValAlaIleAlaPheAlaGluSerArgIleArgAlaGluThrIleAla 344
793 .....TTGGGCGTTCACAGTGCACAAACCGC... 822
345 GluAspValLeuHisAspLeuGlyAlaPheSerLeuThrSerAspSe 361
823 .....CTCTTGGCGTACCGTTTGGGTG 844
361 rGlnAlaMetGlyArgValGlyGluValIleLeuArgThrThrGlnValA 378
845 CGAAAGATATCGCAATTAATCTGCGGCGCAATTGGTT.....GAC 882
378 IaHisArgMetLysValGlnArgIleAlaLeuAlaGluGluThrGlyAsp 394
883 GCAGACAAAC.....CGCGATATTCGCGTTCGTTATGTAACCG 920
395 AsnAspAsnPheArgValLysArgTyrIleAlaLysTyrThrIleAsnTr 411
921 CGCGATTACACAAAGC...GCGCAGCATTAATTTGGACGCTACACCAATC 967
411 oAlaLeuThrHisGlyIleAlaHis.....G 420
968 AGATTTCGTTATTCAGAGAGCGCGACGCAAGAGCTGTTCGGCTGGGT 1017
420 LuValGlySerIleGlyValGlyLysLeuAlaAspLeuValAlaTrp... 435
1018 GCGCGCGACGCGACAAATTAATCTGATCAGCGTACAGCCCTGCGCATTT 1067
436 .....SerProAlaPhePheGlyVal..... 442
1068 CTGGAAGAAACAACTCTTCAGATTTCACGACGACCGTCAACGCTGGCAGC 1117
443 .....LysProAlaThrValIleLysGlyGlyMetI 453
1118 GCCCGCATGGTGGCGATTGGTACTTACGAGCGCGTAAATGCGG 1158
453 leAlaIleAlaProMetGlyAspIleAsnAlaSerIlePro 466

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB pep: US-08-687-645B-5
seq_documentation_block:
; Sequence 5, Application US/08687645B
; Patent No. 5846752
; GENERAL INFORMATION:
; APPLICANT: Robert P. Hausinger
; TITLE OF INVENTION: Mutant Urease and Method
; TITLE OF INVENTION: of Use For Determination
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ian C. McLeod
; STREET: 2190 Commons Parkway
; CITY: Okemos
; STATE: Michigan
; COUNTRY: USA
; ZIP: 48864
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 KB

```

```

MEDIUM TYPE: Storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,645B
FILING DATE: July 26, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Ian C. McLeod
REGISTRATION NUMBER: 20,931
REFERENCE/DOCKET NUMBER: MSU 4.1-309
TELECOMMUNICATION INFORMATION:
TELEPHONE: (517) 347-4100
TELEFAX: (517) 347-4103
TELEX: No. 5846752e
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 567
TYPE: amino acids
STRANDEDNESS: Single
MOLECULE TYPE:
DESCRIPTION: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: N/A
STRAIN: N/A
INDIVIDUAL ISOLATE: N/A
CELL TYPE: N/A
FEATURE:
NAME/KEY: subunit Urec
LOCATION:
IDENTIFICATION METHOD: Sequencing
OTHER INFORMATION: Encoded subunit of mutant
; OTHER INFORMATION: urease
US-08-687-645B-5

alignment_scores:
Quality: 86.50 Length: 247
Ratio: 0.832 Gaps: 12
Percent Similarity: 42.105 Percent Identity: 23.077

alignment_block:
US-09-303-518d-127 x US-08-687-645B-5 ..
Align seg 1/1 to: US-08-687-645B-5 from: 1 to: 567

541 TTGAGCCGTTTGACCGAGCGTAATAATCCATGTGTGAAGCAGCTGGCG 590
|||||
260 LeuAlaAlaIleGlyArgThrIleHisThrPheHisThrGluGlyAl 276
591 AGACGTGCGCTGTGAANAATGCTGCCAACATGGAACACATGATTTGGCG 640
|
276 a.....GlyG 278
641 GCCCGCATCCGGCCGTTTGAGTGGCAGCACATTTCAT...TTCATTGAG 687
|||
278 LYGlyHisAlaProAspIleThrIleAlaHisProAsnIleLeu 294
688 CCGGTCGGTGCACAAACCGCTT...TGGACCATCAATTATGACAGATGT 734
|||
295 ProSerSerThrAsnProThrLeuProTyrThrLeuAsnThrIleAspG 311
735 A.....ATTGCCA 742
311 whisLeuaspMetLeuMetValCysHisHisLeuAspProAspIleAlaG 328

```

```

743 TCAGACGTTGTTTCAGACAGCCGCTGTGACACCGACCGCTGATTGCT 792
      ||||| :|||:||||| :||| :|||
328 LuAspValAlaIphaIaGluSerArgIleArgIleGluThrIleAlaIa 344
      ||||| :|||:||||| :||| :|||
793 .....TTGGGGGTTTCAGTCAAGTCAACAAACACGCG... 822
      ||||| :|||:||||| :||| :|||
345 GluAspValLeuHisAspLeuGlyAlaIphaSerIleuThrSerIleAsp 361
      ||||| :|||:||||| :||| :|||
823 .....CTCTGCGTACCGCTTTTGGGCTG 844
      ||||| :|||:||||| :||| :|||
361 rGlnAlaMetGlyArgValGlyGluValIleLeuArgThrTrpGlnVala 378
      ||||| :|||:||||| :||| :|||
845 CGAAAGTATCGCAATTACTGCGGGCGAATTGTT.....GAC 882
      ||||| :|||:||||| :||| :|||
378 IaHisArgMetLeuValGlnArgGlyAlaLeuAlaGluGluThrGlyAsp 394
      ||||| :|||:||||| :||| :|||
883 GCAGCAAC.....CGCGTATTCGCGTGGTATTTGAACGG 920
      ||||| :|||:||||| :||| :|||
395 AsnAspAsnIleuArgValIleArgIleAlaIleuAlaIleAsnPr 411
      ||||| :|||:||||| :||| :|||
921 CGCGATTCACAGAGC...GCGACGATTTATTTGGACGCTACACATC 967
      ||||| :|||:||||| :||| :|||
411 oAlaLeuThrHisGlyIleAlaHis.....G 420
      ||||| :|||:||||| :||| :|||
968 AGATTTCGTTATCGAGAGAGCGCGACGAAAGAGCTGTCGCGTGGTT 1017
      ||||| :|||:||||| :||| :|||
420 LuValGlySerIleGluValGlyLeuAlaAspLeuValIleTrp... 435
      ||||| :|||:||||| :||| :|||
1018 GCGCGCGACCGCGACAAATFACTCCATCAGCGGTACAGCCCGGCCATT 1067
      ||||| :|||:||||| :||| :|||
436 .....SerProIaIaIphaIheGlyVal..... 442
      ||||| :|||:||||| :||| :|||
1068 CCTGAAAGAAACAACTTCAGTTCAGACAGCGCTCAACGCTGCGACG 1117
      ||||| :|||:||||| :||| :|||
443 .....LysProIaIaIheValIleGlyIleMetI 453
      ||||| :|||:||||| :||| :|||
1118 GCGCGATGTCGCGATGTTGTTACTTACGAGCGCGTAAATGCGG 1158
      ||||| :|||:||||| :||| :|||
453 LeAlaIleAlaIphaMetGlyAspIleAsnAlaSerIlePro 466
      ||||| :|||:||||| :||| :|||
seq_name: /cgn2/6/ptodata/1/1aa/5B_COMB.pep:US-08-456-670B-40
seq_documentation_block:
: Sequence 40, Application US/08456670B
: Patent No. 5932415
: GENERAL INFORMATION:
: APPLICANT: SCHUBERT, PETER
: APPLICANT: NEUMANN, STEFRIED
: APPLICANT: PAMELZIK, MARTINA
: APPLICANT: LINKEWELER, WINFRIED
: APPLICANT: BURGER, CHRISTA
: APPLICANT: HOFMANN, COTTERIED
: APPLICANT: ROBERT, ANDREAS
: APPLICANT: GOEBEL, WERNER
: TITLE OF INVENTION: PROCESSES AND AGENTS FOR DETECTING
: TITLE OF INVENTION: LISTERIAS
: NUMBER OF SEQUENCES: 43
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MILLEN, WHITE, ZELANO & BRANTIGAN, P.C.
: STREET: 2200 CLARENDON BLVD., SUITE 1400
: CITY: ARLINGTON
: STATE: VIRGINIA
: COUNTRY: US
: ZIP: 22201
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/456.670B
: FILING DATE: 01-JUN-1995

```

```

: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/412,227
: FILING DATE: 27-MAR-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/075,248
: FILING DATE: 11-JUN-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: DE 4239567.4
: FILING DATE: 25-NOV-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: DE 421911.4
: FILING DATE: 11-JUN-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: HAWLET-KING, DIANA
: REGISTRATION NUMBER: 33,302
: REFERENCE/DOCKET NUMBER: MERCK 1694D1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-243-6333
: TELEFAX: 703-243-6410
: TELEX: 64191
: INFORMATION FOR SEQ ID NO: 40:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 478 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: ORIGINAL SOURCE:
: ORGANISM: Listeria monocytogenes
: STRAIN: EGD
: US-08-456-670B-40

```

```

alignment_scores:
: Quality: 86.00 Length: 348
: Ratio: 0.494 Gaps: 17
: Percent Similarity: 50.000 Percent Identity: 20.402

```

```

alignment_block:
: US-09-303-518d-127 x US-08-456-670B-40 ..

```

```

Align seg 1/1 to: US-08-456-670B-40 from: 1 to: 478

```

```

145 GATGCGCTCAAAAAGCCCAAGTCTGTTGAACACAAAAGNATCCGGG 194
      ||||| :|||:||||| :||| :|||
50 AspAlaIleLysIleLysAlaAsnAsnIleuThrIleAspIleValProG 66
      ||||| :|||:||||| :||| :|||
195 CGGTGCTTTACCGCGCGCGTTTCAGGCAAAATCGCGCCATCATCGCG 244
      ||||| :|||:||||| :||| :|||
66 yGln.....LysLeuGlnValAsnAsnGluValAlaIaIaIaGluLysT 81
      ||||| :|||:||||| :||| :|||
245 GCGAAAGCGCGCTACTT..... 261
      ||||| :|||:||||| :||| :|||
81 hrcIuIysSerValSerAlaThrTrpLeuAsnValArgThrGlyValaGly 97
      ||||| :|||:||||| :||| :|||
262 .....CACTGCGTCTGATTGCCGTTGAAGCAACGCAAGAAATGAGATT 305
      ||||| :|||:||||| :||| :|||
98 ValAspAsnSerIleIleThrSerIleLysGlyGlyThrLysValThrVa 114
      ||||| :|||:||||| :||| :|||
306 CGAACGCTACGCGCGCGGAGCGTTGGCAAACTTAAGCGCGGANGAANTNN 355
      ||||| :|||:||||| :||| :|||
114 Glu..... 115
      ||||| :|||:||||| :||| :|||
356 GNNCAATCTGATCAATCGCGTTTGTGACATGCGCTGCTGATNCCGTCGG 405
      ||||| :|||:||||| :||| :|||
116 .....ThrThrGluSerAsnGlyTrp..... 122
      ||||| :|||:||||| :||| :|||
406 TTCAGCAAAATCCCTCGCTGCATGCCGAGCGCGTTCGCATTTGTCGA 455
      ||||| :|||:||||| :||| :|||
123 ...HisLysIleThrIleAsnAspGlyLys.....ThrGlyPheValAs 136
      ||||| :|||:||||| :||| :|||
456 TGCATGATGAC...ACCAATCCGCTNCGCGACAGACCTGCTGTGATGATCA 502
      ||||| :|||:||||| :||| :|||

```

```

136 nGlyLysTyrLeuThrAspLysAlaValSerThrProValAlaProThrG 153
503 AAGAGCCGCGCAGCATTTTCAGACAGANGTNGTGTATTGACCGCTTG 552
133 InGUValLysLysGlu.....ThrThrThrGlnGlnAlaAlaProVal 167
553 ACCGAGCCTAAATCCATGTGTGAAGCAGCTGGCCGAGCAGTGGCTC 602
168 AlaGluThrLysThrGluValLysGlnThrGlnAlaThrThrProAl 184
603 TGAATATCTGCGCAACATCGAACAACATGATTTGGCGCCCGCATCCG 652
184 aProLysValAla.....GluThrLysGlu.....ThrProValIleA 197
653 CCGGTGTAGTGGCAGCAGCATTCATTTTCATTTGAGCCGGTGGTCAAC 702
197 spGlnAsnAlaThrThrHis.....AlaValLysSerGly, 208
703 AAAACCGTTTGGACCATCATTTATCAAGANGTAAATTCATCGGACGTTT 752
209 AspThrIleThrPalaLeuSerValLysTyrGlyValSerValGlnAspI 225
753 GTTTCGAACAGCGCGCTCGAACACCGCGCGCTGATTCCT.....T:793
225 emetSerTTPasmsnLeuSerSerSerSerIleTyrValGlyGlnLysL 242
794 TGGGTGGTTCACAGTCACAAACACGACCTCTGC.....829
242 eu.AlaIleLysGlnThrAlaSerThrAlaThrProLysAlaGluValL 258
830 .....GTACCGTTTGGGTGGCAAAAGTATCCGAATTACTGCGGCGCA 872
258 sThrGluAlaProAlaAlaGluLysGlnAlaAlaProValValLysGln 275
873 ATTTGGTACGACGACACACCGCGGATTCCTGGTGGTTCATGACGGCG 922
275 sn.....ThrasnThrAsn.....ThraAla 281
923 CGATTACACAGGCGCGCAGATTAATTGGAGCTACGACATCAATCAGATT 972
282 ThrThrGluLysLysGluThrAlaThrGlnGlnGlnThrAlaProLysAl 298
973 TCCG.....TATCGAAGAGCGCGCAGAAAGACCTGTGGCTGGCT 1016
298 aProThrGluAlaAlaLysProAlaProAlaProSerThrAsnThrAsnA 315
1017 TGGCGCGCAGCGGACAAATCTCATCAGCGGTACGACCTCGGCGCAT 1066
315 laAsnLysThrAsnThrAsnThrAsnThrAsn...AsnThrAsnThrPro 330
1067 TCCTGAAAACAACTCTTCAGATTCAGACAGCGCTCAACG 1108
331 SerLysAsnThrAsnThrAsnSerAsnThrAsnThrAsnThr 344

```

## seq\_name: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:US-08-487-886-2

seq documentation block:

```

Sequence 2, Application US/08487886
Patent No. 5744448
GENERAL INFORMATION:
APPLICANT: Kelton, Christie Ann
APPLICANT: Schweichhardt, Rene Lynn
APPLICANT: Cheng, Shirley Vui Yen
APPLICANT: Nugent, No. 5744488en Patrice
TITLE OF INVENTION: Human Follicle Stimulating
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Stephan P. Williams,
ADDRESSEE: Ares-Serono, Inc.
STREET: Exchange Place, 37th floor
CITY: Boston
STATE: MA

```

```

COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" diskette, 1.44 MB, high density
COMPUTER: IBM PS/2, model 55 SX
OPERATING SYSTEM: MS-DOS version 4.0
SOFTWARE: VAX/VMS Massill via Kermit to IBM MS-DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,886
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/670,085
FILING DATE: 15-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Williams, Stephan P.
REGISTRATION NUMBER: 28546
REFERENCE/DOCKET NUMBER: US/252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 723-1300
TELEFAX: (617) 723-8923
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 695
TYPE: Amino acid
TOPOLOGY: Linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: signal sequence
LOCATION: -17 to -1
IDENTIFICATION METHOD: hydrophobic
FEATURE:
NAME/KEY: putative amino-terminal extracellular domain
LOCATION: 1 to 349
IDENTIFICATION METHOD: similarity with other
IDENTIFICATION METHOD: dimeric glycoprotein receptor extracellular
IDENTIFICATION METHOD: domains, hydrophilic
FEATURE:
NAME/KEY: transmembrane domain
LOCATION: 350 to 613
IDENTIFICATION METHOD: similarity to other G
IDENTIFICATION METHOD: protein-coupled receptor transmembrane domains
FEATURE:
NAME/KEY: putative transmembrane region I
LOCATION: 350 to 370
IDENTIFICATION METHOD: similarity to other G
IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
FEATURE:
NAME/KEY: putative transmembrane region II
LOCATION: 382 to 404
IDENTIFICATION METHOD: similarity to other G
IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
FEATURE:
NAME/KEY: putative transmembrane region III
LOCATION: 427 to 448
IDENTIFICATION METHOD: similarity to other G
IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
FEATURE:
NAME/KEY: putative transmembrane region IV
LOCATION: 469 to 491
IDENTIFICATION METHOD: similarity to other G
IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
FEATURE:
NAME/KEY: putative transmembrane region V
LOCATION: 512 to 533
IDENTIFICATION METHOD: similarity to other G
IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
FEATURE:

```



```

1  MEDIM TYPE: 3.5" diskette, 1.44 MB, high density
2  COMPUTER: IBM PS/2, model 55 SX
3  OPERATING SYSTEM: MS-DOS version 4.0
4  SOFTWARE: VAX/VMS Massill via Kermit to IBM MS-DOS
5  CURRENT APPLICATION DATA:
6  APPLICATION NUMBER: US/08/482,855
7  FILING DATE: 07-JUN-1995
8  CLASSIFICATION: 435
9  PRIOR APPLICATION DATA:
10 APPLICATION NUMBER: 07/670,085
11 FILING DATE: 15-MAR-1991
12 ATTORNEY/AGENT INFORMATION:
13 NAME: Williams, Stephan P.
14 REGISTRATION NUMBER: 28546
15 REFERENCE/DOCKET NUMBER: US/252
16 TELECOMMUNICATION INFORMATION:
17 TELEPHONE: (617) 723-1300
18 TELEFAX: (617) 723-8923
19 INFORMATION FOR SEQ ID NO: 2:
20 SEQUENCE CHARACTERISTICS:
21 LENGTH: 695
22 TYPE: Amino acid
23 TOPOLOGY: Linear
24 MOLECULE TYPE: protein
25 FEATURE:
26 NAME/KEY: signal sequence
27 LOCATION: -17 to -1
28 IDENTIFICATION METHOD: hydrophobic
29 FEATURE:
30 NAME/KEY: putative amino-terminal extracellular domain
31 LOCATION: 1 to 349
32 IDENTIFICATION METHOD: similarity with other
33 IDENTIFICATION METHOD: dimeric glycoprotein receptor extracellular
34 IDENTIFICATION METHOD: domains, hydrophilic
35 FEATURE:
36 NAME/KEY: transmembrane domain
37 LOCATION: 350 to 613
38 IDENTIFICATION METHOD: similarity to other G
39 IDENTIFICATION METHOD: protein-coupled receptor transmembrane domains
40 FEATURE:
41 NAME/KEY: putative transmembrane region I
42 LOCATION: 350 to 370
43 IDENTIFICATION METHOD: similarity to other G
44 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions
45 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
46 FEATURE:
47 NAME/KEY: putative transmembrane region II
48 LOCATION: 382 to 404
49 IDENTIFICATION METHOD: similarity to other G
50 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions
51 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
52 FEATURE:
53 NAME/KEY: putative transmembrane region III
54 LOCATION: 427 to 448
55 IDENTIFICATION METHOD: similarity to other G
56 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions
57 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
58 FEATURE:
59 NAME/KEY: putative transmembrane region IV
60 LOCATION: 469 to 491
61 IDENTIFICATION METHOD: similarity to other G
62 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions
63 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
64 FEATURE:
65 NAME/KEY: putative transmembrane region V
66 LOCATION: 512 to 533
67 IDENTIFICATION METHOD: similarity to other G
68 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
69 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
70 FEATURE:
71 NAME/KEY: putative transmembrane region VI
72 LOCATION: 557 to 580
73 IDENTIFICATION METHOD: similarity to other G

```

```

? IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions
? IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
?
? FEATURE:
?
? NAME/KEY: putative transmembrane region VII
? LOCATION: 592 to 613
? IDENTIFICATION METHOD: similarity to other G
? IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions
? IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
?
? FEATURE:
?
? NAME/KEY: putative carboxy-terminal intracellular
? NAME/KEY: domain
? LOCATION: 614 to 678
?
? DS-08-482-855-2

```

```
alignment_scores:
  Quality: 86.00
  Ratio: 0.480
  Percent Similarity: 46.494
  Length: 3855
  Gaps: 20
  Percent Identity: 20.519
```

alignment\_block:  
US-09-303-518D-127 x US-08-482-855-2

Align seg 1/1 to: US-08-482-855-2 from: 1 to: 695

220 GGCAAAATCGCCGCATCATCGCGGCAAAAGCGCTACTTCAGTCGGT 26  
 ||| ..... ||| ..... ||| ..... ||| .....  
 70 GtysaIseuGtulsYtIleGtIlesterGInAsnAspValIleuGluValII 86  
 270 CGTATTCGCGCTTAAGCGAC ..... GAGCAATCGAGTTCG 307  
 eGluAlasprValIheSerAsnIleProLysIleuHisGluIleKtyleG 103  
 308 AACGC ..... TAGCGCGCGAAGCGTTCGCAAACTTA 339  
 ||| ..... ||| ..... ||| ..... ||| ..... ||| .....  
 103 IuLysAlaAsnAsnIleuYtIleAsnProGluAlaIheGInAsnIleu 119  
 340 AGCGCGCGCAANNNGNCATCTATCGCAACCGCGTTTGAGATGC 389  
 ||| ..... ||| ..... ||| ..... ||| .....  
 120 ProAsnIleuGtInYtIleu ..... LeuIleSerAsn ..... ThrG 121  
 390 GCTCGCGTANCCGTCG ..... TTCAGCAAAATCC 418  
 ||| ..... ||| ..... ||| ..... ||| .....  
 131 YtIleLysIleuProAspValIleHisYtIleHisSerIleuGluVal. 147  
 419 CTGCGGTGAGTATGCGGACCGCTTGCCATCTTCGCAATGGAGATGACACC 468  
 ||| ..... ||| ..... ||| ..... ||| .....  
 148 . LeuIleAspIleGInAsp. . AsnIleAsnIleHisThrIleGluValG 162  
 469 AATCGCGTNCGGCGACACCGCTGTGTGATCAAGAACCGCCGACGAGA 518  
 ||| ..... ||| ..... ||| ..... ||| .....  
 163 AsnSerIheValIcy ..... LeuSe 169  
 519 TTTCAGCGAGANTGTGGTATGAGCGCTTGACCGACCGTAAATCC 568  
 ||| ..... ||| ..... ||| ..... ||| .....  
 169 rIheGluSerValIleIleuYtIleuAsnIuLysAsnGlyIleGInGluIleH 186  
 569 ATGTGTGTAAAGCAGCTGGCGGACAGCGTGCCTGTAAATGCTGCC... 615  
 ||| ..... ||| ..... ||| ..... ||| .....  
 186 IsAsnIcyAlaIheAsnGlyTThrGInIleuAspIleuAsnIleuSerAsp 202  
 616 ..... AACATCGAACA ..... CATGAATTCGCGCG 641  
 ||| ..... ||| ..... ||| ..... ||| .....  
 203 ASGAsnAsnIleuIuIleuProAsnAspValIheHisGlyAlaSerGI 219  
 642 CCGGACATCGCGCGGTGATGAGTGCACATCATCTTCATGAGCGCG 691  
 ||| ..... ||| ..... ||| ..... ||| .....  
 219 yProValIleIleuAspIleSerAlaGhrIleHisSerIleuProSerT 236  
 692 TCGGTCAACAACAACCGTTTGACCATCAATTCAAGATGAATTCGC 741  
 ||| ..... ||| ..... ||| ..... ||| .....  
 236 yGlyIleuIuLysn ..... 240

```

742 ATCGAGCTTTGTTGCAACAGCCCTCTGAC..... 774
   : : : : : : : : : : : : : : : : : :
241 Leuylslyleuargalaargserthrtyrasnleuylsleuproth 257
775 .ACGAGCGCGGTATGCTTGGTGGTTCAGATCAACAAACACGCG 823
   ||| : : : : : : : : : : : : : : : : :
257 rleuglulysleuvalalaleuemetglualaserleuthrtyrpro.Ser 273
824 TCTTCGTCACCGTTTGGGTGCGAAGATATGCAAAATTAATGCGGCGAA 873
   ||| : : : : : : : : : : : : : : : : :
274 Hiscyscysalape.....Alaas 280
874 TTGGTTGACGACAGACCGCGGTATTCGGTTCGGATATGACGCGGC 923
   ||| : : : : : : : : : : : : : : : : :
280 ntrpararglnliserleuHispriole..Cys..Asnlysse 295
924 GATTACACAGAGCGCGACGATTTATTTGGACGCTACCAACATGAG...A 970
   ||| : : : : : : : : : : : : : : : : :
295 rlleleuarglngluvalasptyrmetthrlnthrarglylnarags 312
971 TTTCGGTTATCGACAGAGCGCGACCAAGAGCTGTTCGCTGC..... 1014
   ||| : : : : : : : : : : : : : : : : :
312 erseleualaglulaspasnlguserserTyrserfarglylphespmet 328
1014 ..... 1014
329 thrtYrthrlnphaspTyrAspleucysasnlgulavalaspyalth 345
1015 .GTTCGCGCGACCGCGACCAATATCTCCATCAGCGGTACGACCTCGCGC 1063
   : : : : : : : : : : : : : : : : :
345 rlysserprolysprowalalaphasnprocgslulaspriemetglt 362
1064 ATTCCTGAAAAACAACACTCTTCAGATTTCACGACA.....GCGTCAC 1107
   : : : : : : : : : : : : : : : : :
362 yrasnleuuarlyalleuilertrpheiiserlleuallalethr 378
1108 GGTGGCGACCGCGCGCATGTGCGGATTTGGTACT.....TACGACGCGT 1151
   ||| : : : : : : : : : : : : : : : : :
379 glyasnlllelleuvalleuvalleuintrhrserglntrlylsleuth 395
1152 AATGCGGTAGACATCTCTACCTGCTTTTGGCGGATTTAATGCTG 1201
   : : : : : : : : : : : : : : : : :
395 ryalproargphleuemetCysasnleuallaphaelaaspheucysllieg 412
1202 GC 1203
412 ly 412
seq_name: /cgn2_6/ptodata/1/aa/6B_COMB.pep:US-09-100-193-2
seq_documentation_block:
; Sequence 2, Application US/09100193
; Patent No. 6153729
; GENERAL INFORMATION:
; APPLICANT: GARY S. STEIN ET AL.
; TITLE OF INVENTION: NUCLEAR MATRIX TARGETING PEPTIDES AND USES THEREFORE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/100.193
; FILING DATE:
; CLASSIFICATION:

```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/050,104
FILING DATE: 20-JUNE-1997
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: UMM-024
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 415 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: Internal
US-09-100-193-2

alignment_scores:
Quality: 84.50 Length: 371
Ratio: 0.512 Gaps: 18
Percent Similarity: 44.474 Percent Identity: 19.137

alignment_block:
US-09-303-518D-127 x US-09-100-193-2 ..
Align seg 1/1 to: US-09-100-193-2 from: 1 to: 415

28 CTGCCCATCGCGGGGACGCGGACGAAGTC.....ATTATACGCGGC 71
   ||||| : : : : : : : : : : : :
89 LeuProvalAlaPheLysValAlaLalaLeuGlyAspValProaspGlyTh 105
72 CGTATTACCAAGTCGCGTTCGTTGCGCAAGATATCGCGGTATGCGGC 121
   ||||| : : : : : : : : : : : :
105 rValValThrValMetAlaGlyAsnaspGluAsnTyrSerAla..... 119
122 CCTGATGAAGTCAAGGAAGCGCGATCGCTCAAAAAAGCCAAAGTCTG 171
   : : : : : : : : : : : : : : : : :
120 .....GluLeuArgAsnAlaSerAlaValMetLysAsnGlnValAla 133
172 .TTTGAAGCAAAAAGMATCCGCGC..... 195
   ||||| : : : : : : : : : : : :
134 ProPheAsnAspLeuArgPheValGlyArgSerGlyArgLysSerph 150
196 .....GTGGTTCACCGCGCGCGCGTTCAGCGCAAAATCCCG 232
150 eThrLeuThrIleThrValPheThrAsnProThr.....GlnValAlar 165
233 CCATCATCGCGCGGCAAAAGCGCGTACTTCAATGCGGTGATTCGCGTT 282
   : : : : : : : : : : : : : : : : :
165 hTyrHisArg.....AlaIleLysValThrVal 174
283 GAAGCAACGAAGCAATGCACTGCAACGCTACGCG..... 318
   : : : : : : : : : : : : : : : : :
175 AspGlyProArgGluProArgArgHisArgGlnLysLeuGluAspGlnTh 191
319 .....CCGGAAGCGTTGGCAAACTTAAGCGCGGACGANGAANTMNGN 358
191 rLysProPheProAspArgPheGlyAspLeuGlu..... 202
359 GCAATGTATCAATCCGTTTGTGACTGCGCTGCGGTACGCGTCCGCTTC 408
   ||||| : : : : : : : : : : : :
203 .....ArgLeuArgMetArgValThr 209
409 ACAAATCCCTCGC.....GTGCA 428
   : : : : : : : : : : : : : : : : :
210 ProSerThrProSerProArgGlySerLeuSerThrThrSerHisPhe 226
429 TGGCGAGCGGTGCGCATCTGTCATGCGATGCGATGCAACCAATCGCTG 478
   : : : : : : : : : : : : : : : : :
226 rSerGlnProGlnThrProIleGlnGlyThrSerGluLeuAsnProPhe. 242

```

```
479 CGCGACAGCCCTGTGTGTGATCAAGAAGCCGCGCATGATTTGACAGCA 528
      :|||
243 ..SerAspPro.....ArglnPheValArg 250
529 NCFNNGCTGGTATGACCGCTTGACCGCGTAATC..... 567
      :|||
251 SerPheSerThrLeuSerThrLeuGlnSerArgPheProAspProAr 267
568 .....CATGTCGTAGG 580
267 glleryltyrthrlyalameSerAlaAlaPheProtyrSerAlaThr 284
581 CAGCTGGCGACAGCGTCCGCTGAAATGCTGCACATCGAACA... 627
      :|||
284 roSerGlyThrSerLeuSerLeuSerValAlaGlyIleSerAlaThr 300
628 .....CATGAATTC..... 636
301 SerArgPheHisThrTyrlleuProProtyrProGlyAlaProG1 317
637 .....GGCGCGCCGATCCGCGGTTGTGAGCGACGACATGAT 679
      :|||
317 naSnglnSerGlyProPheGlnAlaasnProSerProtyrHisleutyrt 334
680 TCATTGAGCCGCGTGGTGCACAAACACCGTTGGACCATCATATCA 729
      :|||
334 YrGlyThrSerSerGly.....SerTyrln 342
730 GATGTAATGGCATGCGACGTTGTGTGCAACAGCGCTGCAACCCGA 779
      :|||
343 PheSerMetValAlaGly..... 348
780 GCGGCTGATGCTTGGGTGCTCTCAATGCAACAAACGCGCTTGC 829
      :|||
349 .....SerSerSerGlyAspArgSerProThrArgMetLeuA 362
830 GTACCGTTGGTGGCGAAGTATCGCAATCTGCGCGCAATGCTT 879
      :|||
362 laSerCysThrSerSerAlaIleSer...ValAlaIleGlyAsnLeuMet 377
880 GAGCGACAGACCGCGCTGATTCGGTGTGATGACGCGCGCATTC 929
      :|||
378 AspProSer.....LeuGlyGlyGlnSerAspGlyValGlnAl 390
930 ACAAGGCGCCAC 942
390 aAspGlySerHis 394

seq_name: /cgn2_6/ptodata/1/1aa/PCrus_COMB.pep:PCr-US96-10602-8
seq documentation block:
: Sequence 8, Application PC/TUS9610602
: GENERAL INFORMATION:
: APPLICANT: The General Hospital Corporation
: TITLE OF INVENTION: INHIBITION OF HEPATITIS B REPLICATION
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESS: Fish & Richardson P.C.
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCr/US96/10602
: FILING DATE:
: CLASSIFICATION:
```

```

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/017,814
: FILING DATE: 20-JUN-1995
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Clark, Paul T.
: REGISTRATION NUMBER: 30,162
: REFERENCE/DOCKET NUMBER: 00786/282001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/542-5070
: TELEFAX: 617/542-8906
: TELEX: 200154
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 289 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCr-US96-10602-8

alignment_scores:
  Quality: 84.00      Length: 166
  Ratio: 1.105        Gaps: 13
  Percent Similarity: 45.783  Percent Identity: 29.518

alignment_block:
US-09-303-518d-127 x PCr-US96-10602-8 ..

Align seg 1/1 to: PCr-US96-10602-8 from: 1 to: 289

788 TTGCTTTGGGTGTTCTCAAGTCACAAACGCGCTTCCGACCTT 837
      |||||
100 LeuLeuThrPhe.....HisIleSerCysLeu,Thrp 110
838 TTGGTGGCAAAAGTATTCGCAATTAATCTGCGGCGCAATGTG...TGACGC 884
      |||||
110 heGlyArgGlnThrValIleGlnIuTyrlleuValSerPheGlyValTrpIle 126
885 AGACACCGCGGTGATTCGGGTGCGTATTTGACGCGCGCATTAACAAG 934
      |||||
127 ArgThrProProAlaIleArg.....ProPr 135
935 GCGCGCACGATTAATTTGGAGCGCTACACATCAATGATTCGTTATGAA 984
      :|||
135 oAsnAlaProIleLeuSerThrLeuProGlnThrValValAlaArgArgA 152
985 GAAGCGCGACGAAAGAGCTGTGCGGTGCGCGCGACGCGGACAA 1034
      :|||
152 rg.GlyArgSer.Pro.....ArgArgArgThrP 161
1035 ATACTCCATCACGCGGTACGACCCCTCGGCCATTTCTGTAACAAACTCT 1084
      :|||
161 roSerProAlaArgArgArg.....SerGlnSer 170
1085 TCAAGTTCACGACAGCCGTCACGCTGCGACCGCGCATGTCGCGGAT 1134
      :|||
171 ProArgArgArgArgSerSerThrThrSerThrGlyPro...CysArgTh 186
1135 GGTACTTACGAGCGCGTAATCCGCTAGACATCTGCTGCTTTT 1184
      :|||
186 rCysMetThrThrAla.....GlnGlyThrSerMetTyProSerCysC 201
1185 GCGGATTTAATGCTGC.....GCGATACCGACA 1213
      :|||
201 yScyThrIuysProSerAspGlyAsnCysThrCysIleProIleProSer 217
1214 GCGCGCAAGCATTTGGGT.....GCTTGAATGG 1243
      :|||
218 Ser.....TrpAlaPheGlyLysPheLeuTrpGlnTrp 228

seq_name: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:US-08-467-822-21
```





TITLE OF INVENTION: POLYPEPTIDES  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESS: Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/432,697  
FILING DATE: 02-MAY-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03495, 0137-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ. ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 569 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..569  
OTHER INFORMATION: /note="URE B - FIGURE 3."  
US-08-432-697-21

alignment\_scores:  
Quality: 83.50 Length: 258  
Ratio: 0.898 Gaps: 10  
Percent Similarity: 36.047 Percent Identity: 21.318

alignment\_block:  
US-09-303-518d-127 x US-08-432-697-21 ..

Align seg 1/1 to: US-08-432-697-21 from: 1 to: 569

```
568 CATGTGTGAAGCAGCTGCGCAGACGTC..... 597
||||:||||:||||:||||:||||:||||:||||:
323 HisLeuaspIysSerIleLysgluaspValGlnPhealaaspSerArgIil 339
598 ....CGGTGAAATGCTGCCAATCGAACACATGATGAGTGGGGGCGC 643
||||:||||:||||:||||:||||:||||:||||:
339 eaYpProGlnThrIlealaIalagIuaspIleuHisaspMetGly.... 354
644 CGCATCCGGCCGGTGTGAGTGGCAGCACATTCATTTCATGAGCCGGTC 693
354 ..... 354
694 GGTGCAAAACAAACGTTTGACCATTCATTCAGATGATTCATTCAT 743
||||:||||:||||:||||:||||:||||:||||:
355 .....IlePheSerIleThrSerSerAspSerGlnAlaIle 366
744 CGGACGTTTGTTCGACAGCGCGTGTGAACACGAGCGCGGTGATTCGT 793
||||:||||:||||:||||:||||:||||:||||:
366 tGlyAlaVal.....GlyGluValIleThr..... 374
794 TGGGTGTTCTCAAGTCAACAAACGCGCTTCGTACGCGTTTGGGT 843
||||:||||:||||:||||:||||:||||:||||:
375 .....ArgThrIlePheIleThr 379
```

```
844 GCGAAGATATCGCAATTAATCTGCGGCGAATTCGTT.....GA 881
||||:||||:||||:||||:||||:||||:||||:
380 AlaSpIysasnIySlysgIuPheGlyArgIleuLysgluLysIyS 396
882 CGCAGACAAAC.....CGCGTGAATTCGGGTTCGATTCGACG 919
||||:||||:||||:||||:||||:||||:||||:
396 pasnaspasnPheArgIleLysArgIyIleSerIySlyThrIleasn 413
920 GCGCATTAACAGACGCGCGCATTAATTCGACGATTCACCAATCAG 969
||||:||||:||||:||||:||||:||||:||||:
413 roGlyIlealaHisglYIleSerAspIyValGlySer..... 425
970 ATTTCGCTATCGAAGAGCGCGCAGCAAGAGCTGTTCGCTG..... 1014
||||:||||:||||:||||:||||:||||:||||:
426 .....ValGluValGlySlyThrAlaAspLeuValIleThrSerPr 439
1014 ..... 1014
439 oAlaPhePheGlyIleLysPheAsnMetIleIleLysGlyIlePheIleA 456
1015 .....GTTGCGCGCGACGCG 1029
456 lalauSerGlnMetGlyaspalaAsnAlaserIleProthrProGlnPro 472
1030 GACAAATACATCATCAGCGGTACGACCCCTCGGCCATTCTCTGAAAAACA 1079
||||:||||:||||:||||:||||:||||:||||:
473 ValTyTYr.....ArgIleuMetPheGlyHisIsgIyLysasnly 486
1080 ACTCTTCAGATTCCAGACAGCGCGTCAC..... 1107
||||:||||:||||:||||:||||:||||:||||:
486 sPheaspThrAsnIleThrPheValSerGlnAlaIaTyIrySlaGlyI 503
1108 .....GTTGCGCAGCGCGCATGCTGCCATTCGATTCATTCAC 1143
503 lElysgIuLysgluLysgluLysaspArgalaIaIaProValIySasnly 519
1144 GAGCGCGTAAATCCCGCTAGACATC 1167
520 ArgAsnIleThrLysLysAspIleu 527

seq_name: /cgn2_6/prodata/1/1aa/5B_COMB.pep:US-08-466-248-21
seq_documentation_block:
; Sequence 21, Application US/08466248
; Patent No. 6258359
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerdam, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thibierge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSER: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,248
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/447,177
FILING DATE: 19-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/432,697
FILING DATE: 02-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03495, 0137-02000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 569 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Protein
LOCATION: 1..569
OTHER INFORMATION: /note="URE B - FIGURE 3."
US-08-466-248-21

```

```

alignment_scores:
Quality: 83.50 Length: 258
Ratio: 0.898 Gaps: 10
Percent Similarity: 36.047 Percent Identity: 21.318

```

```

alignment_block:
US-09-303-518D-127 x US-08-466-248-21 ..

```

```

Align seg 1/1 to: US-08-466-248-21 from: 1 to: 569

```

```

568 CATGTGCTAAGCAGCTGGCCGACGCTG..... 597
      |||||  |||||
333 HisleuAspLysSerIleLysGlnAspValGlnPheAlaAspSerArgI1 339
598 ....CCGCTGAAATGCTGCCAATGCAACATGATGATTCGGCGCC 643
      |||||  |||||  |||||  |||||
339 eArgProGlnThrIleAlaAlaGluAspGlnLeuHisAspMetGly... 354
644 CGCATCCGCGCGTTGAGTGCGACGACATTCATTGAGCCGCTC 693
354 ..... 354
694 GGTGCAACAAACCGTTTGACCATCAATTATCAAGTGAATTCGCAT 743
      |||||  |||||  |||||  |||||
355 .....IlePheSerIleThrSerSerAspSerGlnAlaIle 366
744 CGAGCTTTGTTTGCACAGCGCGCTGACACCGCGCGTGAATTCGCT 793
      |||||  |||||  |||||  |||||
366 tGlyArgVal.....GlyGluValIleThr..... 374
794 TGGGTGTTCTCAAGTCAACAAACCGCTTCGCTACCGTTTGGGT 843
      |||||  |||||
375 .....ArgThrTrpGlnThr 379
844 GCGAAGTATCGCAATTAATCTCGGCGCAATTCGTT.....GA 881
      |||||  |||||  |||||  |||||
380 AlaAspLysAsnLysGlnPheGlyArgLeuLysGluLysGlyAs 396
882 CCGACAGCAAC.....CCGCTGATTCGCGTTCGGTATTCGACG 919
      |||||  |||||  |||||  |||||
396 pAsnAspAsnPheArgIleLysArgTyrIleSerLysTyrThrIleAsn 413
920 GCGCATTAACACAGCGCGACGATTATTGGAGCGCTACACAAATCAG 969
      |||||  |||||  |||||  |||||

```

```

413 roGlyIleAlaHisGlyIleSerAspTyrValGlySer..... 425
970 ATTCCGTTATGAGAGAGCGCGACCAAGAGCTGTTCGCTG..... 1014
      |||||  |||||  |||||  |||||
426 .....ValGluValGlyLysTyrAlaAspLeuValLeuTrpSerPr 439
1014 ..... 1014
439 AlaPhePheGlyIleLysProAsnMetIleIleLysGlyGlyPheIleA 456
1015 .....GTTCGCCGCGACCGC 1029
456 ILeuSerGlnMetGlyAspAlaAsnAlaSerIleProThrProGlnPro 472
1030 GACAAATACCTCCATCAGCGGTAGACCCCGCGCATTCCTGAAAAACA 1079
      |||||  |||||  |||||  |||||
473 ValTyrTyr.....ArgGluMetPheGlyHisHisGlyLysAsnLys 486
1080 ACTCTCAAGTTCACGACGACCGCTCAAC..... 1107
      |||||  |||||
486 sPheAspThrAsnIleThrPheValSerGlnAlaIleTyrLysAlaGlyI 503
1108 .....GGTGGCAGCCGCGCATGGTGGCATTCGCTACTTAC 1143
503 IeLysGluGluLeuGlyLeuAspArgAlaIleAlaProValLysAsnCy 519
1144 GAGCGCGTAAATGCGCGTACGATC 1167
      |||||  |||||
520 ArgAsnIleThrLysLysAspLeu 527

```

```

seq_name: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:US-09-199-637A-351

```

```

seq_documentation_block:

```

```

Sequence 351 Application US/09199637A
Patent No. 6353411
GENERAL INFORMATION:
APPLICANT: Ausubel, Frederick
APPLICANT: Goodman, Howard M.
APPLICANT: Rahme, Laurence G.
APPLICANT: Mahajan-Miklos, Shalina
APPLICANT: Tan, Man-wah
APPLICANT: Cao, Hui
APPLICANT: Tsongalis, John
TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
FILE REFERENCE: 00786/361002
CURRENT FILING DATE: 1998-11-25
PRIOR APPLICATION NUMBER: US/09/199,637A
PRIOR FILING DATE: 1997-11-25
NUMBER OF SEQ ID NOS: 437
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 351
TYPE: PR1
ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-351

```

```

alignment_scores:
Quality: 83.00 Length: 277
Ratio: 0.741 Gaps: 14
Percent Similarity: 40.433 Percent Identity: 25.271

```

```

alignment_block:

```

```

US-09-303-518D-127 x US-09-199-637A-351 ..
Align seg 1/1 to: US-09-199-637A-351 from: 1 to: 447

```

```

15 AAAAGCTTAACCTGCCCATTCGCGGCGACAGC.....GACACAGTCA 58
      |||||  |||||  |||||  |||||
210 ArgArgProAlaAspGlyHisAlaGlyGlnThrAlaProGlyAspProG1 226
      |||||  |||||  |||||  |||||

```



```
589 GCAGAGTCGCCG...TCTGAAATGCTGCCAATCGAACAACATGATTT 635
::|||::||| ::::: :::::
1864 ThrAspAlaProGlnAlaArgAspValThrValAlaProLeuLeuArgAl 1880
::|||::||| ::::: :::::
636 CGCGGCGCCGCGATCCG.....CCCGGTTTGAGTGGCAGCCACA 673
||| |||::||| ::::: :::::
1880 aglyArgProGluProGluThrLeuAlaIaGlyLeuAlaThrAlaHisV 1897
764 TTCAT.....TTCATGAGCCGCGTCCGTGCAAC 702
::||| ::::: ||| ||| :::::
1897 aHisGlyAlaProLeuAspArgAlaSerPhePheProAspGlyArgArg 1913
703 AAACCGTTTGAGCATCATATTATCAAGATGTAAATGCCATCGACGTTT 752
::||| ::::: :::::
1914 ThrAspLeuProThrTyrAlaPheArg.....GluHisTyr 1926
753 GTTGGCAACAGCGCGCTGTGAACCGGCGCGTGATGCTTTGGGTGGTT 802
::||| ::::: ::::: ||| ||| :::::
1926 rTTPLeuThrProGluAlaArgThrAsp...AlaArgAlaLeuGlyPheA 1942
803 CTCAGTCAACAACCAACCGCCTCTTGCGTACCGTT...TTGGGTGCGAA 849
::|||::||| ::::: ||| ||| :::::
1942 spProAlaArgHisProLeuLeuThrThrValGluValAlaGlyGly 1958
850 GATCGCAATTTACTCGCGGCGAATTTGTTGACGACACAACCGCGTGAT 899
::|||::||| ::::: ||| ||| :::::
1959 AspGlyValLeuLeuThrGlyArgLeuSerLeuThrAspGlnProTyrPle 1975
900 TTCGCGTTGCGTATGTAACGCGCGATTTACACAAGCGCGCGCATTTAT 949
::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1975 uAlaAspHisMetValAsnGlyAlaValLeuLeuProAlaThrAlaPheL 1992
950 TG 951
||
1992 eu 1992
```

